UC Davis

UC Davis Previously Published Works

Title

Distinctive Patterns of Flavonoid Biosynthesis in Roots and Nodules of *Datisca glomerata* and *Medicago spp.* Revealed by Metabolomic and Gene Expression Profiles.

Permalink

https://escholarship.org/uc/item/9q89x7bs

Authors

Gifford, Isaac
Battenberg, Kai
Vaniya, Arpana
et al.

Publication Date

2018

DOI

10.3389/fpls.2018.01463

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

frontiers in Plant Science

ORIGINAL RESEARCH
published: 10 October 2018
doi: 10.3389/fpls.2018.01463

Distinctive Patterns of Flavonoid Biosynthesis in Roots and Nodules of *Datisca glomerata* and *Medicago spp.* Revealed by Metabolomic and Gene Expression Profiles

Isaac Gifford[1], Kai Battenberg[1†], Arpana Vaniya[2†], Alex Wilson[1], Li Tian[1], Oliver Fiehn[2] and Alison M. Berry[1]*

[1] Department of Plant Sciences, University of California, Davis, Davis, CA, United States, [2] West Coast Metabolomics Center, University of California, Davis, Davis, CA, United States

OPEN ACCESS

Edited by:
Ulrike Mathesius, Australian National University, Australia

Reviewed by:
Maciej Stobiecki, Institute of Bioorganic Chemistry (PAS), Poland
Zhentian Lei, University of Missouri, United States

***Correspondence:**
Isaac Gifford
isgifford@ucdavis.edu

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Evolution and Development, a section of the journal Frontiers in Plant Science

Received: *08 May 2018*
Accepted: *14 September 2018*
Published: *10 October 2018*

Citation:
Gifford I, Battenberg K, Vaniya A, Wilson A, Tian L, Fiehn O and Berry AM (2018) Distinctive Patterns of Flavonoid Biosynthesis in Roots and Nodules of Datisca glomerata and Medicago spp. Revealed by Metabolomic and Gene Expression Profiles. Front. Plant Sci. 9:1463. doi: 10.3389/fpls.2018.01463

Plants within the Nitrogen-fixing Clade (NFC) of Angiosperms form root nodule symbioses with nitrogen-fixing bacteria. Actinorhizal plants (in Cucurbitales, Fagales, Rosales) form symbioses with the actinobacteria *Frankia* while legumes (Fabales) form symbioses with proteobacterial rhizobia. Flavonoids, secondary metabolites of the phenylpropanoid pathway, have been shown to play major roles in legume root nodule symbioses: as signal molecules that in turn trigger rhizobial nodulation initiation signals and acting as polar auxin transport inhibitors, enabling a key step in nodule organogenesis. To explore a potentially broader role for flavonoids in root nodule symbioses across the NFC, we combined metabolomic and transcriptomic analyses of roots and nodules of the actinorhizal host *Datisca glomerata* and legumes of the genus *Medicago.* Patterns of biosynthetic pathways were inferred from flavonoid metabolite profiles and phenylpropanoid gene expression patterns in the two hosts to identify similarities and differences. Similar classes of flavonoids were represented in both hosts, and an increase in flavonoids generally in the nodules was observed, with differences in flavonoids prominent in each host. While both hosts produced derivatives of naringenin, the metabolite profile in *D. glomerata* indicated an emphasis on the pinocembrin biosynthetic pathway, and an abundance of flavonols with potential roles in symbiosis. Additionally, the gene expression profile indicated a decrease in expression in the lignin/monolignol pathway. In *Medicago sativa*, by contrast, isoflavonoids were highly abundant featuring more diverse and derived isoflavonoids than *D. glomerata.* Gene expression patterns supported these differences in metabolic pathways, especially evident in a difference in expression of cinnamic acid 4-hydroxylase (C4H), which was expressed at substantially lower levels in *D. glomerata* than in a *Medicago truncatula* transcriptome where it was highly expressed. C4H is a major rate-limiting step in phenylpropanoid biosynthesis that separates the pinocembrin pathway from the lignin/monolignol and naringenin-based flavonoid branches. Shikimate *O*-hydroxycinnamoyltransferase, the link between

flavonoid biosynthesis and the lignin/monolignol pathway, was also expressed at much lower levels in *D. glomerata* than in *M. truncatula*. Our results indicate (a) a likely major role for flavonoids in actinorhizal nodules, and (b) differences in metabolic flux in flavonoid and phenylpropanoid biosynthesis between the different hosts in symbiosis.

Keywords: flavonoid, root nodule, symbiosis, actinorhizal, legume, metabolome profile, gene expression profile, phenylpropanoid

INTRODUCTION

Root nodule symbioses (RNS) develop as symbiotic associations between nitrogen-fixing bacteria and certain host plants, resulting in the formation of the root nodule, a specialized organ for nitrogen fixation and assimilation. The root nodule provides a number of functions in RNS, primarily serving as a site for the exchange of carbon and energy-containing molecules from the host for nitrogen-containing molecules from the microsymbiont, and also as an environment to help regulate oxygen concentration to protect the nitrogenase enzyme complex. The bacteria capable of establishing these symbioses fall into two distantly related groups: the proteobacterial rhizobia and the actinobacterial genus *Frankia*. The host plants, on the other hand, all belong to a single clade of angiosperms known as the Nitrogen-fixing Clade (NFC) (Soltis et al., 1995), consisting of the order Fabales (nodulated by rhizobia) and three orders that include the actinorhizal plants, Cucurbitales, Fagales, and Rosales (nodulated by *Frankia*).

The establishment of RNS involves a signal–mediated recognition interaction between host and microsymbiont within the rhizosphere, followed by the entry of the microsymbiont into root cells, and ultimately by nodule organogenesis (Gage, 2004). Early stages of organogenesis involve the division of cortical cells and cell expansion during invasion by the microsymbiont, followed by nodule organogenesis and maturation of nitrogen-fixing symbiotic tissue. During the maturation phase, cells within the developing nodule undergo endoreduplication, increasing in volume and becoming more transcriptionally active to promote symbiotic interactions (Vinardell et al., 2003). In both the legume and actinorhizal symbioses several of the initial steps in the internal signaling pathway leading to nodule establishment are conserved. Initial signaling interactions are activated via the Common Symbiotic Pathway, a set of genes shared with the more ancient arbuscular mycorrhizal symbioses (Oldroyd, 2013), indicating a shared evolutionary origin within the NFC (Markmann and Parniske, 2009; Battenberg et al., 2018), followed by a RNS-specific gene expression cascade (Oldroyd, 2013).

Flavonoids are ubiquitous secondary metabolites synthesized by the phenylpropanoid pathway. The flavonoid pathway is one of two major branches in plant phenylpropanoids, the other being monolignol/lignin biosynthesis, and is responsible for producing a wide range of metabolites fundamental for plant structure and function and plant–organism interactions including symbiotic signaling in RNS and nodule organogenesis and development. Plant flavonoids are also key molecules in pigmentation and signaling for pollinator attraction, herbivore or pathogen deterrence, reduction of damage from reactive oxygen species, UV light protection, and regulation of development (Shirley, 1996; Buer et al., 2010). The reactions linking the flavonoid and monolignol/lignin branches of the pathway and the interconversions among flavonoid classes are illustrated in **Figure 1**.

In the establishment of root nodule symbiosis in many legume genera, flavonoids produced by the host are recognized by rhizobia in the rhizosphere, primarily through the receptor-transcription factor NodD. This, in turn, triggers the expression of the other *nod* genes (*nodA, nodB, nodC*), which synthesize a lipochitooligosaccharide molecule, the Nod factor, which is secreted by the rhizobia (Long, 1996), that in turn triggers host cellular responses leading to root-nodule development (Oldroyd, 2013). A wide range of flavonoid molecules, both aglycones and glycosides, has been identified as nodulation signals in legume symbioses (Peck et al., 2006). To date, no molecule similar to Nod factor has been identified in the *Frankia*-actinorhizal symbioses; however, genomes of some members of the Cluster 2 group of *Frankia* contain homologs of the rhizobial *nodABC* genes that are expressed during symbiosis (Persson et al., 2015; Nguyen et al., 2016), suggesting that a Nod factor may play a role in at least some actinorhizal symbioses. Cluster 2 *Frankia* genomes do not contain any identified homologs of *nodD*, leaving the mechanism of induction of transcription unknown, and the role of flavonoids in actinorhizal-*Frankia* signaling to be determined.

After the initial signaling steps in nodulation, flavonoids play a continuing role in legume nodule development. Flavonoids are known to bind to and inhibit auxin transporters, leading to a disruption of polar auxin transport (Wasson et al., 2006). The accumulation of auxin within certain cortical cells in the root triggers cell division and proliferation, which, in turn, leads to the localized induction of the nodule (Mathesius et al., 1998). Recent studies have suggested that the production of flavonoids during nodule organogenesis is itself a response to increased cytokinin production following the perception of the symbiotic Nod factor (Mathesius et al., 1998). Similar effects in the actinorhizal symbioses have been far less studied; however, it has been shown that inhibition of auxin gradients with an auxin influx inhibitor in the actinorhizal host *Casuarina glauca* led to decreased nodulation, suggesting a similar role for auxin in actinorhizal symbioses (Péret et al., 2007; Champion et al., 2015). Additionally, flavonoids have been suggested to play a role in triggering endoreduplication through DNA breaks resulting in anaphase arrest (Cantero et al., 2006).

In this study, flavonoids from the metabolomes of roots and nodules of the actinorhizal host *Datisca glomerata* were compared with those of the legume *Medicago sativa*. Additionally, the metabolome results were compared with available transcriptomes of *D. glomerata* and *Medicago*

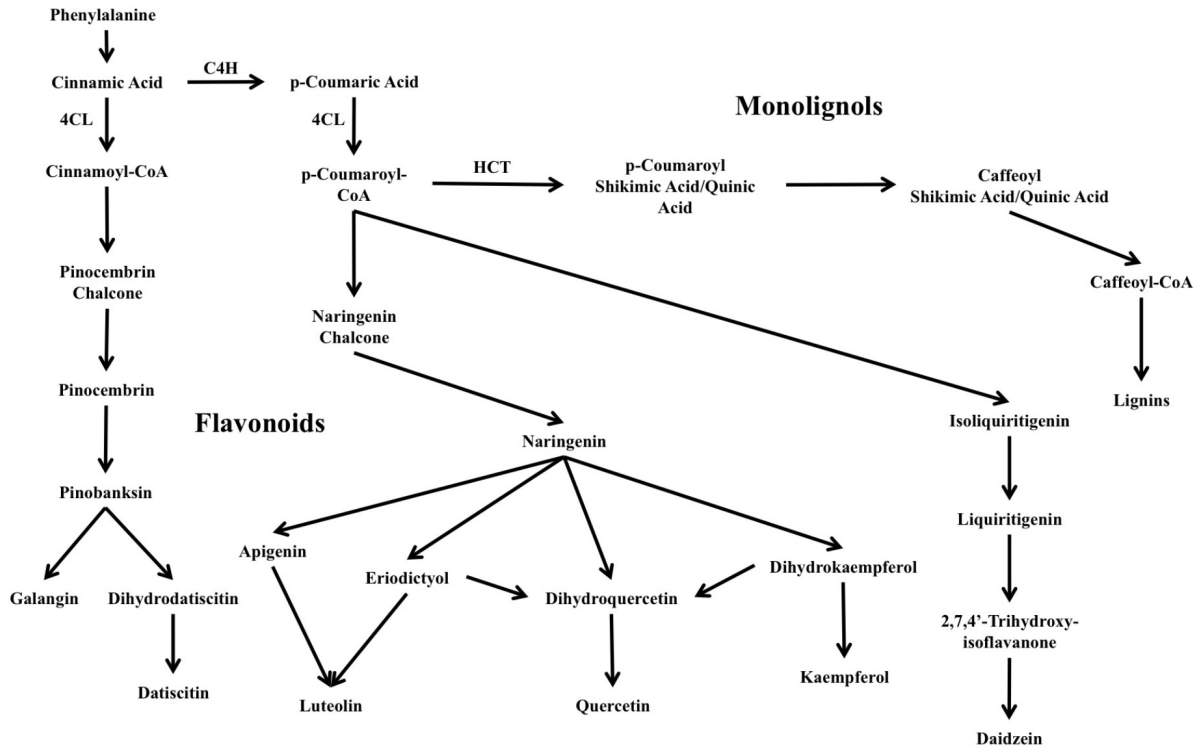


FIGURE 1 | Overview of major flavonoid biosynthesis pathways in *Datisca glomerata*. Also shown is the lignin/monolignol pathway (upper-right). C4H: cinnamic acid 4-hydroxylase. 4CL: 4-coumarate CoA-ligase. HCT: shikimate *O*-hydroxycinnamoyltransferase.

truncatula (Roux et al., 2014; Battenberg et al., 2018). Both hosts were found to synthesize phenylpropanoid derivatives of the flavonoid branch in several different categories including flavones, flavanones, and isoflavonoids, but with different apparent patterns of metabolic flux.

MATERIALS AND METHODS

Growth Conditions and Nodule Sampling

Datisca glomerata seeds were collected from wild plants growing in Gates Canyon, Vacaville, CA, United States, germinated, grown, and inoculated in a greenhouse at University of California, Davis, under conditions as described in Battenberg et al. (2018). The seedlings were inoculated with crushed *Ceanothus thyrsiflorus* nodules containing *Frankia* originally sampled in Sagehen Experimental Forest (Truckee, CA, United States). Until inoculation, one-half-strength Hoagland's solution with nitrogen (Hoagland and Arnon, 1950) was applied weekly. After inoculation, one-half-strength Hoagland's solution without nitrogen was applied weekly.

Uninoculated roots, inoculated roots, and root nodules were collected for analysis from four individual plants per treatment. Inoculated roots and nodules were collected from the same plants; both were harvested 100 days after inoculation. Uninoculated roots were collected from the same plants sampled for inoculated roots and nodules, prior to inoculation. Sampling methods are described in detail in Battenberg et al. (2018). Collected samples were flash frozen in liquid nitrogen and stored at −80°C until use. For detailed information on samples collected, see **Supplementary Table S1**. Root and nodule samples from individual plants were ground in liquid nitrogen in a mortar and pestle, prior to extraction.

Mature individual *M. sativa* plants were collected from field plantings at the Russell Ranch Sustainable Agriculture Facility, University of California, Davis, and maintained in a greenhouse at University of California, Davis. Roots and root nodules were collected and sampled from four individual plants per treatment, as described above.

Flavonoid Extraction

Root nodules and inoculated roots of *M. sativa* and *Datisca glomerata* were extracted using 80:20 MeOH/H_2O. 40 mg of samples were extracted with 2000 μL of cold solvent. Samples were then mixed for 10 s using Mini Vortexer (VWR, Radnor, PA, United States). Samples were then centrifuged for 5 min at 14,000 relative centrifugal force (RCF) using an Eppendorf Centrifuge 5415D (Hauppaugee, NY, United States). After removing the supernatant, samples were dried using a Labconco CentriVap Concentrator (Kansas City, MO, United States). Dried samples were resuspended in 110 μL of 10:90 ACN/H_2O with 1 μg/mL 12-(cyclohexylcarbamoylamino)dodecanoic acid (CUDA) for LC-MS/MS analysis.

Metabolomic Analysis

For metabolomic analysis of flavonoids and related molecules, a comparison was made between root nodules and inoculated roots collected from the same plants. Chromatography was performed using a Thermo Vanquish UHPLC instrument, a Phenomenex Kinetex C18 column (100 × 2.1 mm, 1.7 μm) with a KrudKatcher Ultra HPLC in-line filter (0.5 μm Depth Filter × 0.004 in ID). The mobile phases were H_2O with 0.1% acetic acid (A) and ACN with 0.1% acetic acid (B). Gradient elution was performed at a flow rate of 0.5 mL/min under the following program: from 0 to 10 min B changed linearly from 10 to 90%, held at 90% B for 2.50 min, returned to 10% B over the next 2.50 min, and held at 10% B for equilibration over 5 min. The column temperature was kept at 45°C. The LC method was modified from Ma et al. (2016). MS/MS data were acquired on a high resolution Thermo Q Exactive HF mass spectrometer in positive electrospray ionization (ESI) mode under the following operating parameters: sheath gas flow rate at 60, auxiliary gas flow rate at 25, sweep gas flow rate at 2, spray voltage at 3.60 kV, capillary temperature at 300°C, S-lens RF level at 50, and auxiliary gas heater temperature at 370°C. Mass spectral data were collected using full scan MS1 and data-dependent MS/MS. Full scan MS1 had the following parameters: scan range from a *m/z* 150–2000 with the resolution set to 120,000, AGC target set to 1×10^6, and maximum ion injection set to 300 ms. Data-dependent MS2 had the following parameters; scan range from *m/z* 150–2000 with the resolution set to 15,000, AGC target set to 1×10^5, maximum injection time set to 50 ms, loop count set to 3, and TopN set to trigger the top-3 most abundant ions, with an isolation window of 1.0 *m/z*, and Higher Energy Collisional Dissociation (HCD) was conducted using three normalized collision energies; 35, 45, and 65%. The observed MS/MS spectra have an HCD collision energy of 48.33%, spectra from the three normalized collision energies are automatically averaged. The injection volume for each sample was 2 μL.

In metabolomics compounds are routinely identified by data processing tools which match MS/MS spectra against mass spectral reference libraries and use cheminformatics to provide spectral interpretation (Neumann and Bocker, 2010; Dunn et al., 2012; Cajka and Fiehn, 2015). Here we have used MS-DIAL software version 2.82 (Tsugawa et al., 2015) was used to process the raw data and metabolites were reported using a 0% peak count filter to keep all detected features. MS-DIAL was used for data deconvolution, peak alignment, and compound identification by searching mass spectral refrence libraries. Compound identifications were made based on an in-house accurate mass and retention time (m/z-RT) library created from the QC reference standard mix and the following tandem mass spectral libraries; MassBank, ReSpect, MetaboBASE, HMDB, GNPS, NIST 17 MS/MS, FAHFA, LipidBlast, and iTree MS/MS only. The tandem mass spectral libraries were downloaded in an msp format from MassBank of North America (MoNA) which was later used in MS-DIAL. MS-FLO (Mass Spectral Feature List Optimizer) was used as post processing tool to optimize the feature list from MS-DIAL to remove duplicate and isotopic features and identifiy ions adduct (DeFelice et al., 2017).

After reduction, annotations were also labeled with Metabolomics Standards Intitaive (MSI) levels and mass error (mDa) to provide confidence in each annotation. Level 1 is the highest level of identification. It is described as using two or more orthogonal data from an authentic standard. Level 2 is when only one set of reference data from an authentic standard has been used, for example, either using an in-house accurate m/z-RT library or using mass spectral library search for MS/MS matching. Level 3, is similar to Level 2 where a match can be made with either a m/z-RT library or a MS2 library, but the match lacks high accuracy. Lastly, Level 4 indicates those metabolites that are unknown (Sumner et al., 2007; Schymanski et al., 2014). The observed MS spectra of identified flavonoid compounds in *D. glomerata* and *M. sativa* are shown in **Supplementary Figures S1**, **S2**, respectively, as head-to-tail comparisons of experimental and reference MS/MS spectra.

Metabolome and Phenylpropanoid Pathway Analysis

Flavonoid molecules detected by LCMS were annotated by their International Chemical Identifier (InChIKey) and separated into subclasses of flavonoids, isoflavonoids, flavones, flavonols, and anthocyanins by ClassyFire (Feunang et al., 2016). For each plant, the average peak height of flavonoids in each plant was calculated and molecules with average peak heights above the mean were considered highly abundant. Significant differences between roots and nodules were identified with two-tailed Welch's *t*-tests. A comparison of the overall proportions of flavonoids annotated by class in the metabolomes of *D. glomerata* and *M. sativa* was performed with a chi-square test. T-tests and chi-square tests were performed in R using a significance level of $p < 0.05$.

For molecules with multiple annotated isotopes, only the dominant isotope was used, identified by the highest peak height across all samples. Phenylpropanoid biosynthesis pathways were obtained from KEGG (Kanehisa et al., 2016). Maps used included: Flavonoid Biosynthesis, Flavone and Flavonol Biosynthesis, Anthocyanin Biosynthesis, Isoflavonoid Biosynthesis, and Phenylpropanoid Biosynthesis. Within each class molecules that were significantly different between root and nodule LC-MS samples were grouped by their structural similarity using molecular structures acquired from PubChem (Kim et al., 2016) to reference molecules that included: eriodictyol, naringenin, liquiritigenin, daidzein, genistein, glycitein, formononetin, kaempferol quercetin, luteolin, apigenin, cyanidin, pinocembrin, pinobanksin, galangin, and chrysin. In cases where an enzyme for synthesizing a particular molecule could not be identified, putative pathways were inferred, placing molecules together in groups if their chemical structures shared diagnostic structures of the reference molecules including: a 2C–3C carbon double bond, a 3 carbon hydroxyl, 3′ or 4′ hydroxyls, or a 2 or 3 carbon benzene ring.

Abundant Compound Verification by HPLC, UV Absorption and LC-MS

To investigate the effect of *Frankia* interactions with *D. glomerata* on flavonoid production uninoculated roots, roots inoculated

with *Frankia*, and nodules of *D. glomerata* were collected for a second analysis of selected abundant flavonoids. The ground tissue (100 mg) was extracted in 300 μL of 80% methanol, with incubation in an ultrasonic water bath for 20 min at 30°C. The extract was then centrifuged twice for 10 min each at 17,000 × *g*. The supernatant was transferred to an HPLC vial; 10 μL of the supernatant was injected on a reverse phase HPLC and analyzed as previously described (Knollenberg et al., 2018). Major metabolites (i.e., abundant HPLC peaks) that exhibited differential accumulation among roots and nodules were collected and analyzed by mass spectrometry (MS) in negative mode and MS/MS using an established method (Ono et al., 2016). The flavonoid metabolites were tentatively identified based on their retention times, UV absorption spectra, as well as MS and MS/MS data, taking into consideration phenolic metabolites previously reported to accumulate in roots of Datiscaceae (Bohm, 1988). In addition, authentic standards of kaempferol, luteolin, and genistein (Sigma Aldrich, St. Louis, MO, United States) were analyzed in parallel with the *D. glomerata*.

One-way analysis of variation (ANOVA) followed by Tukey's HSD test were performed on the metabolite data using JMP (SAS Institute, Cary, NC, United States).

Transcriptome Analysis of *D. glomerata* and *M. truncatula* Phenylpropanoid Pathways

Transcript annotations from Battenberg et al. (2018) were used for *D. glomerata* and *M. truncatula* (Roux et al., 2014). The transcriptome of *M. truncatula* was chosen because of its depth of coverage and comparable stage of nodulation. Enzyme Commission (EC) number annotations were made with InterProScan v5.21 (Jones et al., 2014) and Trinotate v3.0.1 (Haas et al., 2013). Transcripts annotated with Enzyme Commission (EC) numbers belonging to the KEGG phenylpropanoid biosynthetic pathways listed above were identified in each transcriptome. Expression fold changes between nodules and roots (Log-scale) were used to generate heat maps in Microsoft Excel. To determine genes important for symbiosis and compare relative expression levels of genes between nodules of the two hosts, transcripts in the 90th percentile or above ranked by transcripts per million (TPM) in the full transcriptomes were considered "highly expressed," and transcripts below the 50th percentile were considered "low expression." Statistical significance of expression level differences between *D. glomerata* and *M. truncatula* nodules were determined with one-sample *t*-tests comparing the percentile ranks of the most highly expressed transcript for each gene in the *D. glomerata* transcriptome with the most highly expressed transcript from *M. truncatula* ($p < 0.05$).

RESULTS

Metabolomics Analysis Revealed Abundant and Diverse Flavonoid Accumulation in *D. glomerata* and *M. sativa*

In total, 384 compounds from *D. glomerata* roots and nodules were initially annotated as flavonoids, however, only 60

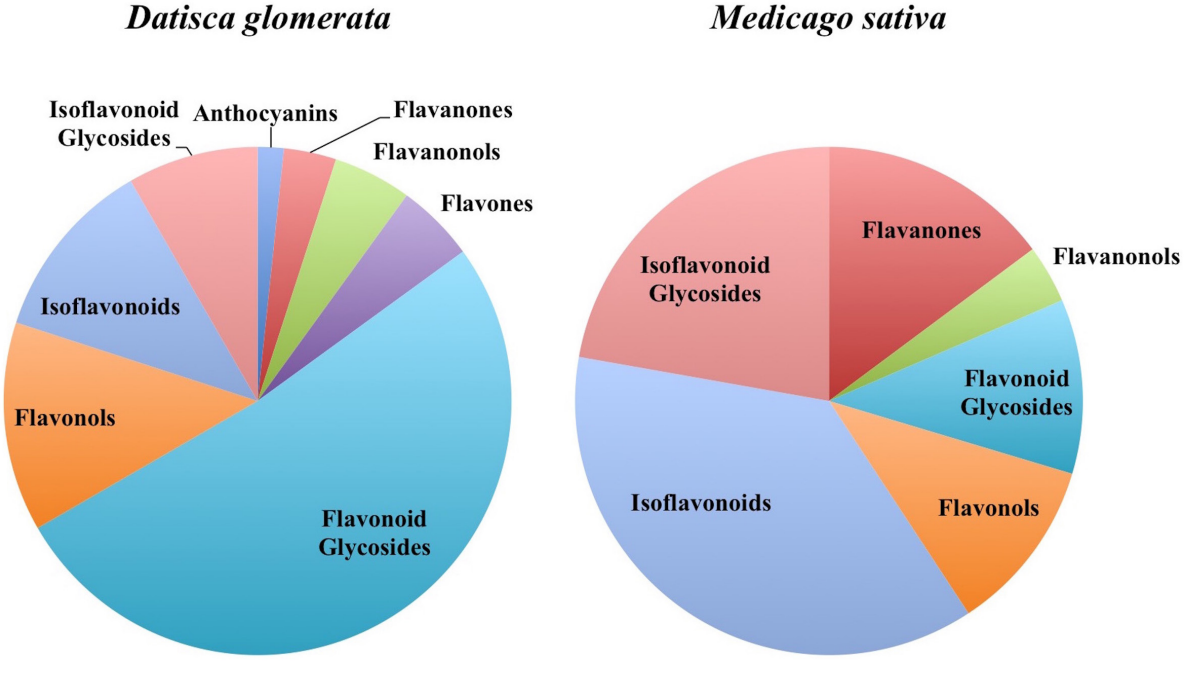


FIGURE 2 | Distribution of flavonoids in *D. glomerata* and *M. sativa* roots and nodules by class.

metabolites were matched against spectra from an MS/MS library at high levels of identification (MSI Level 1, 2, or 3) and the rest were reclassified as unknowns (**Supplementary Table S2**). The relative distributions of annotated flavonoids by class in *D. glomerata* and *M. sativa* are presented in **Figure 2** and listed in **Figures 3**, **4** for *D. glomerata* and *M. sativa*, respectively. Of the 60 annotated flavonoids in *D. glomerata* 24 were aglycones and 36 were glycosides, the majority of which were flavonols (**Figure 3**). In *M. sativa* 281 compounds were initially annotated as flavonoids but only 27 compounds met the MSI Level 3 or better, of which nine were glycosylated. *M. sativa* root and nodule flavonoids were predominantly isoflavonoids (**Figure 4**). These differences in flavonoid distribution between *D. glomerata* and *M. sativa* were strongly significant ($p < 5 \times 10^{-8}$, **Supplementary Table S3**). With the exception of glycitin in *M. sativa* nodules all flavonoids identified were detected in both the roots and nodules of their respective plant (**Figures 3**, **4**).

Ten flavonoids were highly abundant in nodules of *D. glomerata*. Abundant aglycones included isoquercitin (*m/z* 465.1021), quercetin (*m/z* 303.0495), galangin (*m/z* 271.0597), pinocembrin (*m/z* 257.0805), datiscetin (*m/z* 287.0545), cirsiliol (*m/z* 331.0807), and daidzein (*m/z* 255.0648), while the abundant glycosylated flavonoids included the datiscetin-derivative datiscin (*m/z* 595.1652) and two putative derivatives of kaempferol (*m/z* 449.1072 and *m/z* 639.1916) (**Figure 3**). The relative abundance of these flavonoids in *D. glomerata* roots and nodules, determined by *t*-tests, was not significantly different between roots and nodules.

In *M. sativa* nodules, the highly abundant flavonoids were exclusively isoflavonoids, including formononetin (*m/z* 269.0813), the most abundant, and its derivative the second most abundant flavonoid medicarpin (*m/z* 271.0968), as well as coumestrol (*m/z* 269.0449) and the genistein-derivative prunetin (*m/z* 285.0762) (**Figure 4**). Due to sample variability, statistically significant fold changes were not obtained for the majority of the identified *M. sativa* flavonoids between nodules and roots.

Differential Accumulation of Flavonoids in *D. glomerata* Roots and Nodules

Flavonoids

In *D. glomerata* nodules, naringenin and pinocembrin (*m/z* 257.0805), both of which are flavanones from which other flavonoid classes are derived, were significantly increased over roots (**Figure 3**). Dihydrokaempferol (*m/z* 271.0597), an aglycone derivative of naringenin that is an intermediate in the synthesis of flavonols from naringenin, was also significantly increased in the nodule.

Flavonols and Flavones

As noted above, the most abundant flavonoids in *D. glomerata* nodules were the flavonols datiscetin and datiscin, as well as galangin, all of which are derivatives of pinocembrin (**Figure 3**). Other flavonols, including kaempferol and quercetin, were found to have several known and putative derivatives significantly more abundant in the nodules than roots as well. Three putative kaempferol glycosides were highly abundant in *D. glomerata* nodules: astragalin (*m/z* 449.1071), kaempferol 7-*O*-glucoside (*m/z* 449.1072), and demethoxycentaureidin 7-*O*-rutinoside (*m/z* 639.1916); and quercetin and its derivative isoquercetin (*m/z* 465.1021) were highly abundant in *D. glomerata* nodules.

No flavones, putative derivatives of apigenin and luteolin, were found to be significantly different between *D. glomerata* roots and nodules (**Figure 3**).

Isoflavonoids

Daidzein (*m/z* 255.0648) was one of the most abundant flavonoids in *D. glomerata* nodules and one of its derivatives, 6′-*O*-acetyldaidzin (*m/z* 459.1279) was significantly increased in nodules over roots (**Figure 3**). Additionally, genistein (*m/z* 271.0597), belonging to a separate isoflavonoid pathway, was also significantly more abundant in nodules than roots.

Anthocyanins

Only one anthocyanin was annotated from *D. glomerata*: peonidin 3-*O*-rutinoside (*m/z* 463.1229) and it was neither abundant nor significantly different in the nodule (**Figure 2**).

Rutinose Glycosides Accumulated in *D. glomerata* Nodules

A considerable number of rutinose glycosides were identified in the flavonoids of *D. glomerata* roots and nodules, across several of the flavonoid classes, including datiscin (*m/z* 595.1652), one of the most abundant flavonoids identified (**Figure 3**). Other rutinose glycosides detected included rutin (quercetin-3-*O*-rutinoside, *m/z* 611.1600), narirutin (naringenin-7-*O*-rutinoside, *m/z* 419.1331), peonidin-3-*O*-rutinoside (*m/z* 463.1229), demethoxycentaureidin-7-*O*-rutinoside (*m/z* 639.1916), and kaempferol-3-*O*-rutinoside (*m/z* 595.1653). None of these were significantly different between roots and nodules, e.g., more abundant in nodules than roots. Of the rutinose glycosides, only narirutin was identified in *M. sativa* roots or nodules (**Figure 3**).

Nodulation Enhanced Flavonoids in Inoculated *D. glomerata* Roots and Nodules

To examine how nodulation may influence flavonoid metabolites, we compared abundant metabolite composition of non-inoculated roots, inoculated roots, and nodules. Six metabolites (peaks 1–6) showed significantly greater accumulation in nodules in comparison to non-inoculated roots (**Figure 5**). Five (peaks 2–6) showed significantly greater accumulation in nodules than in the inoculated roots (**Figure 5**). In addition, two of the metabolites examined (peaks 2 and 3) showed significantly greater accumulation in inoculated roots when compared to non-inoculated roots. These metabolites were tentatively identified as structurally related flavonoids and flavonoid glycosides based on their retention times, UV absorption spectra, as well as MS and MS/MS data. The most abundant compound (peak 3) in all samples was tentatively identified as datiscetin (3,5,7,2′-Tetrahydroxyflavone) with other compounds (peaks 1, 2, and 5) potentially representing methylated or glycosylated derivatives. Peaks 4 and 6 were tentatively identified as kaempferol and galangin, respectively.

Datisca glomerata

Annotation	*m/z* (Da)	Theoretical *m/z* (Da)	Mass error (mDa)	Adduct	Exact Mass (Da)	Dot Product	Reverse Dot Product	Retention Time	InChIKey	Log Root Average Peak Height (Arbitrary Units)	Log Nodule Average Peak Height (Arbitrary Units)	Log Fold Change (Root vs Nodule)	p-value (Root vs Nodule)	Class
Peonidin 3-*O*-rutinoside	463.1229	463.1235	0.6	[Cat-C6H10O4]+	609.1814	718.3713	872.0741	2.20	ONQVTPMFYSRRLL-HUJUZFAXSA-O	6.40	5.80	-0.60	0.063	Anthocyanin
Maritimein	449.1072	449.1079	0.7	[M+H]+	448.1006	975.1226	692.7123	2.72	SYRURBPRFQUVQS-RHEJLWEFSA-N	7.27	7.59	0.32	0.092	Aurone Glycoside
4-Deoxyphloridzin	443.1307	443.1312	0.5	[M+Na]+	420.1420	869.4292	871.8372	2.91	NLGUKXQDDTZCDG-QNDFHXLGSA-N	6.70	6.97	0.27	0.018	Chalcone Glycoside
Naringin dihydrochalcone	437.1431	437.1443	1.2	[M+H-C6H10O4]+	582.1949	742.5112	663.6321	0.60	CWBZAESOUBENAP-QVNVHUMTSA-N	5.89	5.09	-0.80	0.075	Chalcone Glycoside
Naringenin	273.0754	273.0758	0.4	[M+H]+	272.0685	382.3394	920.6653	2.27	FTVWIRXFELQLPI-UHFFFAOYSA-N	6.45	7.04	0.59	0.005	Flavanone
Pinocembrin	257.0805	257.0808	0.3	[M+H]+	256.0735	878.6736	926.7058	2.57	URFCJEUYXNAHFI-UHFFFAOYSA-N	7.03	7.39	0.36	0.029	Flavanone
Naringin	419.1331	419.1337	0.6	[M+H-C6H10O5]+	580.1792	895.2618	881.4257	1.91	HXTPHSYLYXVTHC-AJHDJQFQSA-N	6.28	5.75	-0.54	0.201	Flavanone Glycoside
Naringenin 4',7-dimethylether	167.0338	167.0339	0.1	[M+H-C9H10O]+	300.0998	915.0733	932.6581	1.49	CKEXCBVNKRHAMX-HNNXBMFYSA-N	6.42	6.36	-0.05	0.657	Flavanone, O-methylated
3,5,7,3',4'-Pentahydroxyflavanone	287.0546	287.0556	1.0	[M+H-H2O]+	304.0583	949.9904	940.8458	0.49	CXQWRCVTCMQVQX-UHFFFAOYSA-N	5.69	6.69	1.00	0.032	Flavanonol
Fustin	243.0649	243.0652	0.3	[M+H-CH2O2]+	288.0634	589.035	822.2004	0.80	FNUPUYFWZXZMIE-LSDHHAIUSA-N	6.73	7.70	0.98	0.003	Flavanonol
Dihydrokaempferol	271.0597	271.0596	0.1	[M+H-H2O]+	288.0634	831.9346	812.5168	0.80	PADQINQHPQKXNL-LSDHHAIUSA-N	6.80	7.73	0.93	0.003	Flavanonol
Chrysin	255.0647	255.0652	0.5	[M+H]+	254.0579	-1	-1	4.00	RTIXKCRFFJGDPG-UHFFFAOYSA-N	4.98	5.52	0.53	0.204	Flavone
Luteolin	287.0546	287.0550	0.4	[M+H]+	286.0477	583.1896	878.9459	9.55	IQPNAANSBPBGFQ-UHFFFAOYSA-N	6.11	6.24	0.14	0.356	Flavone
6-Hydroxyflavone	239.07	239.0703	0.3	[M+H]+	238.0630	-1	-1	3.53	GPZYYYGYCRFPBU-UHFFFAOYSA-N	5.44	5.52	0.08	0.591	Flavone
Apigenin 7-glucoside	433.1124	433.1130	0.6	[M+H]+	432.1057	904.0886	856.6778	2.36	KMOUJOKENFFTPU-QNDFHXLGSA-N	6.90	7.33	0.44	0.204	Flavone Glycoside
Luteolin-4'-*O*-glucoside	449.1073	449.1079	0.6	[M+H]+	448.1006	867.0875	747.9448	0.78	UHNXUSWGOJMEFO-QNDFHXLGSA-N	7.00	7.28	0.29	0.055	Flavone Glycoside
Vitexin-2''-rhamnoside	579.1703	579.1709	0.6	[M+H]+	578.1636	837.1746	967.3208	1.56	LYGPBZVKGHHTIE-UHFFFAOYSA-N	5.40	5.34	-0.07	0.857	Flavone Glycoside
Luteolin-7,3'-di-*O*-glucuside	611.1600	611.1607	0.7	[M+H]+	610.1534	645.8632	747.78	2.14	BISZYPSIZGKOFA-IPOZFMEPSA-N	5.96	5.74	-0.22	0.374	Flavone Glycoside
Apigenin 7-*O*-neohesperidoside	579.1703	579.1709	0.6	[M+H]+	578.1636	797.7675	817.7665	0.44	FKIYLTVJPDLUDL-WAEXOFCTSA-N	6.14	5.48	-0.66	0.115	Flavone Glycoside
Pectolinarin	623.1965	623.1971	0.6	[M+H]+	622.1898	662.697	879.4876	2.79	DUXQKCCELUKXOE-CBBZIXHGSA-N	7.56	6.89	-0.67	0.248	Flavone Glycoside
Isorhoifolin	579.1703	579.1709	0.6	[M+H]+	578.1636	814.5816	692.4609	2.21	FKIYLTVJPDLUDL-SLNHTJRHSA-N	7.53	6.74	-0.79	0.213	Flavone Glycoside
2-(3,4-dihydroxyphenyl)-5,7-dihydroxy-6-methoxychromen-4-one	317.0652	317.0656	0.4	[M+H]+	316.0583	618.3651	853.3499	2.90	FIHSFRSDKWJKJ-UHFFFAOYSA-N	6.13	6.20	0.07	0.645	Flavone, O-methylated
Cirsiliol	331.0807	331.0812	0.5	[M+H]+	330.0739	781.0229	806.0188	3.80	IMEYGBIXGJLUIS-UHFFFAOYSA-N	7.95	7.91	-0.04	0.813	Flavone, O-methylated
Acacetin	285.0754	285.0758	0.4	[M+H]+	284.0685	849.429	905.3773	5.44	DANYIYRPLIHOCZ-UHFFFAOYSA-N	7.68	5.63	-2.05	0.254	Flavone, O-methylated
Rhamnetin	317.0652	317.0656	0.4	[M+H]+	316.0583	853.1801	655.4908	2.48	JGUZGNYPMHHYRK-UHFFFAOYSA-N	6.69	7.14	0.44	0.006	Flavonol
Myricetin	319.0444	319.0449	0.5	[M+H]+	318.0376	806.1461	477.5073	1.19	IKMDFBPHZNJCSN-UHFFFAOYSA-N	6.71	6.84	0.13	0.809	Flavonol
Quercetin	303.0495	303.0500	0.5	[M+H]+	302.0427	973.4973	990.4039	1.57	REFJWTPEDVJJIY-UHFFFAOYSA-N	7.78	7.90	0.12	0.577	Flavonol
Datiscetin	287.0545	287.0550	0.5	[M+H]+	286.0477	988.1427	988.7355	2.95	WCNLFPKXBGWWDS-UHFFFAOYSA-N	9.28	9.37	0.09	0.570	Flavonol
Quercetin 3'-methyl ether	317.0651	317.0656	0.5	[M+H]+	316.0583	958.5301	967.8045	1.98	IZQSVPBOUDKVDZ-UHFFFAOYSA-N	6.60	6.46	-0.13	0.559	Flavonol
Galangin	271.0597	271.0601	0.4	[M+H]+	270.0528	953.0203	871.7371	4.16	VCCRNZQBSJXYJD-UHFFFAOYSA-N	7.84	7.68	-0.16	0.592	Flavonol
Morin	303.0495	303.0500	0.5	[M+H]+	302.0427	960.5632	960.0779	1.31	YXOLAZRVSSWPFT-UHFFFAOYSA-N	6.60	6.44	-0.16	0.639	Flavonol
3,7-Dihydroxy-3',4'-dimethoxyflavone	315.0858	315.0863	0.5	[M+H]+	314.0790	698.1826	883.3394	5.06	LTSJOKAWSCMBC-UHFFFAOYSA-N	7.72	6.26	-1.47	0.235	Flavonol
Engeletin	289.0703	289.0707	0.4	[M+H-C6H10O4]+	434.1213	717.6806	881.7886	1.48	VQUPQWGKORWZII-WDPYGAQVSA-N	5.74	6.24	0.51	0.009	Flavonol Glycoside
Spiraeoside	303.0495	303.0500	0.5	[M+H-C6H10O5]+	464.0955	826.7088	911.2938	1.06	OIUBYZLTFSLSBY-HMGRVEAOSA-N	7.37	7.85	0.48	0.004	Flavonol Glycoside
2-(3,4-dihydroxyphenyl)-5-hydroxy-7-methoxy-3-[3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxychromen-4-one	479.118	479.1184	0.4	[M+H]+	478.1111	731.8151	736.6822	2.06	PHEWILLIAJUBQE-UHFFFAOYSA-N	6.67	6.82	0.15	0.288	Flavonol Glycoside
5-hydroxy-2-[2-hydroxy-3-[(2S,3R,4S,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxyphenyl]-7,8-dimethoxychromen-4-one	493.1335	493.1341	0.6	[M+H]+	492.1268	795.8863	793.5306	2.76	JCUIPEIMZRLNKQ-BSTKLLGTSA-N	7.49	7.64	0.15	0.206	Flavonol Glycoside
Quercetin 3-glucoside	487.0842 465.1024	487.0847 465.1028	0.5 0.4	[M+Na]+ [M+H]+	464.0955	851.1539	890.3217	0.68 0.68	OVSQVDMCBVZWGM-QSOFNFLRSA-N	7.03	7.00	-0.04	0.824	Flavonol Glycoside
Astragalin	449.1071	449.1079	0.8	[M+H]+	448.1006	928.9191	852.9584	2.12	JPUKWEQWGBDDQB-QSOFNFLRSA-N	8.31	8.27	-0.04	0.656	Flavonol Glycoside
Isoquercitin	487.084 465.1021	487.0847 465.1028	0.7 0.7	[M+Na]+ [M+H]+	464.0955	957.9615	934.7609	1.55 1.55	OVSQVDMCBVZWGM-LQSBFMDOSA-N	8.12	8.03	-0.09	0.678	Flavonol Glycoside
Isorhamnetin 3-*O*-glucoside	479.118	479.1184	0.4	[M+H]+	478.1111	931.8452	928.2596	1.98	CQLRUIIRRZYHHS-LFXZADKFSA-N	6.89	6.73	-0.16	0.510	Flavonol Glycoside
3-[(2S,4S,6R)-6-[[(2R,3R,4R,5S,6S)-3,5-dihydroxy-6-methyl-4-[(2S,3R,4S,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxyoxan-2-yl]oxymethyl]-3,4,5-trihydroxyoxan-2-yl]oxy-5,7-dihydroxy-2-(4-hydroxyphenyl)chromen-4-one	757.2179	757.2186	0.7	[M+H]+	756.2113	704.0867	702.8189	1.87	MFIXKYXSBNIMPX-PMLYRHJJSA-N	6.96	6.33	-0.63	0.105	Flavonol Glycoside
3-[(2S,3R,4S,5S,6R)-4,5-dihydroxy-6-(hydroxymethyl)-3-[(2S,3R,4S,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxyoxan-2-yl]oxy-2-(3,4-dihydroxyphenyl)-5-hydroxy-7-methoxychromen-4-one	641.1708 663.1526	641.1712 663.1531	0.4 0.5	[M+H]+ [M+Na]+	640.1639	682.3683 -1.0	754.4238 -1.0	1.17	CEZKIFXYWPTANH-HGVYTWBRSA-N	5.86	5.18	-0.68	0.095	Flavonol Glycoside
Kaempferol 7-*O*-glucoside	449.1072	449.1079	0.7	[M+H]+	448.1006	857.2097	865.2679	1.72	YPWHZCPMOQGCDQ-HMGRVEAOSA-N	7.91	7.14	-0.77	0.106	Flavonol Glycoside
3-[(2S,3R,4S,5S,6R)-4,5-dihydroxy-6-(hydroxymethyl)-3-[(2S,3R,4S,5R,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxyoxan-2-yl]oxy-5-hydroxy-2-(4-hydroxyphenyl)-7-methoxychromen-4-one	625.1758 647.1577	625.1763 647.1582	0.5 0.5	[M+H]+ [M+Na]+	624.1690	653.0514 -1.0	749.4342 -1.0	2.54	PNBMEXOTFKNHLQ-PVZBESAUSA-N	6.81	6.02	-0.79	0.050	Flavonol Glycoside
3-[(2S,3R,4S,5S,6R)-4,5-dihydroxy-3-[3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxy-6-[[(2R,3R,4R,5R,6S)-3,4,5-trihydroxy-6-methyloxan-2-yl]oxymethyl]oxan-2-yl]oxy-5,7-dihydroxy-2-(4-hydroxyphenyl)chromen-4-one	757.2181	757.2186	0.5	[M+H]+	756.2113	669.0611	696.2686	1.64	VNLOLXSJMINBIS-CTLJHIMSSA-N	6.52	5.72	-0.80	0.069	Flavonol Glycoside
Kaempferol 3-*O*-rutinoside	595.1653	595.1657	0.4	[M+H]+	594.1584	795.3095	831.0267	1.96	RTATXGUCZHCSNG-QHWHWDPRSA-N	6.20	5.40	-0.80	0.134	Flavonol Glycoside
Datiscin	595.1652 617.1471	595.1657 617.1476	0.5 0.5	[M+H]+ [M+Na]+	594.1584	912.4401 861.3495	855.3838 930.0329	1.720 1.730	BJCTXDEJUWVJC-QHWHWDPRSA-N	8.46	7.45	-1.00	0.112	Flavonol Glycoside
Isorhamnetin 3-neohesperidoside	625.1759	625.1763	0.4	[M+H]+	624.1690	871.5287	845.5419	1.93	QHLKSZBFIJJREC-SPSUIZEHSA-N	6.53	5.45	-1.08	0.131	Flavonol Glycoside
Rutin	611.1600	611.1607	0.7	[M+H]+	610.1534	-1	-1	1.51	IKGXIBQEEMLURG-NVPNHPEKSA-N	7.07	5.99	-1.08	0.134	Flavonol Glycoside
Robinin	741.2225	741.2237	1.2	[M+H]+	740.2164	635.501	930.6442	1.71	PEFASEPMJYRQBW-HKWQTAEVSA-N	6.38	4.82	-1.56	0.140	Flavonol Glycoside
Demethoxyencentaureidin 7-*O*-rutinoside	639.1916	639.1920	0.4	[M+H]+	638.1847	648.081	863.0599	2.34	PYPKJBUJNZMSTH-UHFFFAOYSA-N	7.76	5.80	-1.97	0.163	Flavonol Glycoside
3-[4,5-dihydroxy-3-[(2R,3R,4R,5R,6S)-3,4,5-trihydroxy-6-methyloxan-2-yl]oxy-6-[[(2R,3R,4R,5R,6S)-3,4,5-trihydroxy-6-methyloxan-2-yl]oxymethyl]oxan-2-yl]oxy-5-hydroxy-2-(4-hydroxyphenyl)-7-methoxychromen-4-one	755.2383	755.2394	1.1	[M+H]+	754.2321	713.8873	696.2553	2.61	JGDLCWFCXJLERP-ZSTRHOSHSA-N	6.20	3.96	-2.24	0.206	Flavonol Glycoside
Genistein	271.0597	271.0601	0.4	[M+H]+	270.0528	951.3821	953.8713	1.49	TZBJGXHYKVUXJN-UHFFFAOYSA-N	6.63	7.44	0.81	0.004	Isoflavonoid
Daidzein	255.0648	255.0652	0.4	[M+H]+	254.0579	751.7332	676.6323	1.34	ZQSJRDFPHDXIC-UHFFFAOYSA-N	7.85	7.64	-0.21	0.285	Isoflavonoid
6''-*O*-Acetyldaidzin	459.1279	459.1286	0.7	[M+H]+	458.1213	593.218	750	2.28	ZMOZJTDOTOZVRT-DODNOZFWSA-N	4.92	5.84	0.92	0.002	Isoflavonoid Glycoside
Genistin	433.1124	433.113	0.6	[M+H]+	432.1057	-1	-1	1.54	ZCOLJUOHXJRHDI-IJFYFLOXSA-N	6.33	6.76	0.43	0.139	Isoflavonoid Glycoside
Tectoridin	463.1228	463.1235	0.7	[M+H]+	462.1162	610.8192	817.1731	2.99	CNOURESJATUGPN-UDEBZQQRSA-N	7.68	7.66	-0.02	0.884	Isoflavonoid Glycoside
Sophoricoside	433.1123	433.1130	0.7	[M+H]+	432.1057	876.6987	901.6234	2.21	ISQRJFLLIDGZEP-CMWLGVBASA-N	6.99	6.22	-0.77	0.218	Isoflavonoid Glycoside
Glycitin	447.128	447.1286	0.6	[M+H]+	446.1213	818.9316	758.4519	3.03	OZBAVEKZGSOMOJ-MIUGBVLSSA-N	6.68	5.71	-0.97	0.284	Isoflavonoid Glycoside
Pratensein	301.0703	301.0707	0.4	[M+H]+	300.0634	782.0549	753.2568	2.53	FPIOBTBNRZPWJW-UHFFFAOYSA-N	6.60	6.41	-0.19	0.176	Isoflavonoid, O-methylated

Scale	Log Root	Log Nodule	Log Fold Change
	10.00	10.00	2.00
	5.00	5.00	1.00
	0.00	0.00	0.00
			-1.00
			-2.00

FIGURE 3 | Flavonoids identified in *D. glomerata* roots and nodules. All annotations conform to MSI level 2. For each flavonoid the abundance in both the roots and nodules is given as the log average peak height of 10 samples and presented as a heat-map. The names of molecules considered highly abundant are highlighted in blue. Significant changes in abundance between roots and nodules are highlighted in yellow.

The accumulation of these metabolites in roots and nodules are consistent with previous reported flavonoid profiles of *D. glomerata* (Bohm, 1988). These metabolite identifications were further supported by the observations that the retention time and UV absorption spectra for peak 3 (datiscetin; $[M\text{-}H]^-$ *m/z* 285.0369) was inconsistent with an authentic standard of luteolin (*m/z* 286.2390) that has the same MS fragmentation pattern as datiscetin. Similarly, the retention time and UV absorption spectra for peak 6 (galangin; $[M\text{-}H]^-$ *m/z* 269.0434) were inconsistent with authentic standards of apigenin (*m/z* 270.2400) and genistein (*m/z* 270.2400) that share the same MS fragmentation pattern as galangin. The data for peak 4 was consistent with an authentic standard of kaempferol analyzed in parallel.

Medicago sativa

Annotation	*m/z* (Da)	Theoretical *m/z* (Da)	Mass error (mDa)	Adduct	Exact Mass (Da)	Dot Product	Reverse Dot Product	Retention Time	InChIKey	Root Average Peak Height (Arbitrary Units)	Nodule Average Peak Height (Arbitrary Units)	Fold Change (Root vs Nodule)	p-value (Root vs Nodule)	Class
Isoliquiritin	419.1343	419.1337	0.6	[M+H]+	418.1264	944.5847	943.5583	2.18	YNWXJFQOCHMPCK-LXGDFETPSA-N	5.89	6.03	0.14	0.627	Chalcone Glycoside
7-Hydroxyflavanone	241.0863	241.0859	0.4	[M+H]+	240.0786	981.1453	988.3962	4.56	SWAJPHCXKPCPQZ-UHFFFAOYSA-N	5.56	6.90	1.34	0.339	Flavanone
Liquiritigenin	257.0812	257.0808	0.4	[M+H]+	256.0735	905.5894	927.9195	3.31	FURUXTVZLHCCNA-UHFFFAOYSA-N	6.60	7.31	0.71	0.296	Flavanone
Pinocembrin	257.0813	257.0808	0.5	[M+H]+	256.0735	787.3966	907.5052	3.92	URFCJEUYXNAHFI-UHFFFAOYSA-N	5.63	6.11	0.48	0.495	Flavanone
Naringenin	273.0762	273.0758	0.4	[M+H]+	272.0685	-1	-1	2.72	FTVWIRXFELQLPI-ZDUSSCGKSA-N	5.73	6.13	0.39	0.408	Flavanone
Narirutin	419.1332	419.1337	0.5	[M+H-C6H10O5]+	580.1792	864.0896	811.4224	1.78	HXTFHSYLYXVTHC-AJHDJQPGSA-N	4.67	4.77	0.10	0.683	Flavanone Glycoside
Dihydrokaempferol	271.0606	271.0596	1.0	[M+H-H2O]+	288.0634	840.4518	822.2188	0.58	PADQINQHPQKXNL-LSDHHAIUSA-N	6.97	6.72	-0.25	0.058	Flavanonol
Naringin	581.1872	581.1865	0.7	[M+H]+	580.1792	-1	-1	1.85	DFPMSGMNTNDNHN-CSIAVLANSA-N	3.59	4.21	0.62	0.293	Flavone Glycoside
Quercetin 3'-methyl ether	317.0661	317.0656	0.5	[M+H]+	316.0583	924.9923	975.8453	3.1	IZQSVPBOUDKVDZ-UHFFFAOYSA-N	4.03	4.26	0.23	0.575	Flavonol
Datiscetin	287.0555	287.0550	0.5	[M+H]+	286.0477	911.2479	889.0267	2.95	WCNLFPKXBGWWDS-UHFFFAOYSA-N	4.31	4.29	-0.02	0.956	Flavonol
3-Hydroxy-3'-methoxyflavone	269.0812	269.0808	0.4	[M+H]+	268.0735	783.5383	962.6492	4.56	GYLGASXCHFNKHD-UHFFFAOYSA-N	7.25	5.63	-1.62	0.322	Flavonol
Coumestrol	269.0449	269.0445	0.4	[M+H]+	268.0372	972.4412	959.3188	2.9	ZZIALNLLNHEQPJ-UHFFFAOYSA-N	6.69	7.66	0.97	0.165	Isoflavonoid
Daidzein	255.0656	255.0652	0.4	[M+H]+	254.0579	-1	-1	2.18	ZQSIJRDFPHDXIC-UHFFFAOYSA-N	7.23	7.06	-0.17	0.493	Isoflavonoid
Medicarpin	271.0968	271.0965	0.3	[M+H]+	270.0892	940.9747	825.0352	3.75	NSRJSISNDPOJOP-BBRMVZONSA-N	8.33	8.03	-0.31	0.228	Isoflavonoid
Genistein	271.0605	271.0601	0.4	[M+H]+	270.0528	-1	-1	2.79	TZBJGXHYKVUXJN-UHFFFAOYSA-N	5.35	4.97	-0.38	0.104	Isoflavonoid
Prunetin	285.0762	285.0758	0.4	[M+H]+	284.0685	742.4944	674.7249	2.93	KQMVAGISDHMXJJ-UHFFFAOYSA-N	8.06	7.66	-0.41	0.101	Isoflavonoid
Glycitein	285.0762	285.0758	0.4	[M+H]+	284.0685	-1	-1	2.36	DXYUAIFZCFRPTH-UHFFFAOYSA-N	5.54	5.13	-0.41	0.063	Isoflavonoid
Glycitin	447.1293	447.1286	0.7	[M+H]+	446.1213	-1	-1	1.09	OZBAVEKZGSOMOJ-MIUGBVLSSA-N	4.13	0.00	0.00	0.066	Isoflavonoid Glycoside
Sissotrine	469.1113 447.1294	469.1105 447.1286	0.8 0.8	[M+Na]+ [M+H]+	446.1213	751.821	788.7471	1.72	LFEUICHQZGNOHD-UHFFFAOYSA-N	6.31	6.12	-0.18	0.271	Isoflavonoid Glycoside
Ononin	453.1163 431.1343	453.1156 431.1337	0.7 0.6	[M+Na]+ [M+H]+	430.1264	965.7795	839.9406	2.16	MGJLSBDCWOSMHL-MIUGBVLSSA-N	7.80	7.36	-0.44	0.036	Isoflavonoid Glycoside
6''-*O*-Malonylglycitin	533.1301	533.1290	1.1	[M+H]+	532.1217	836.0235	702.8371	2.31	OWMHCYFEUPHFB-GOZZSVHWSA-N	6.79	5.98	-0.81	0.134	Isoflavonoid Glycoside
Formononetin 7-*O*-glucoside-6''-*O*-malonate	517.1349	517.1341	0.8	[M+H]+	516.1268	864.2612	773.4846	2.67	RDTAGQKYPGLCBK-GOZZSVHWSA-N	8.09	7.16	-0.93	0.030	Isoflavonoid Glycoside
6''-*O*-Malonyldaidzin	503.1194	503.1184	1.0	[M+H]+	502.1111	864.9066	739.4436	1.86	MTXMHWSVSZKYBT-ASDZUOGYSA-N	6.24	5.07	-1.18	0.076	Isoflavonoid Glycoside
Formononetin	269.0813	269.0808	0.5	[M+H]+	268.0735	914.7	967.7452	3.38	HKQYGTCOTHHOMP-UHFFFAOYSA-N	9.40	8.95	-0.45	0.038	Isoflavonoid, O-methylated
3',4'-Dimethoxy-7-hydroxyisoflavone	299.0919	299.0914	0.5	[M+H]+	298.0841	836.9082	916.3702	3	VFZIJLPRJAQGFO-UHFFFAOYSA-N	6.54	5.90	-0.64	0.083	Isoflavonoid, O-methylated
6,4'-Dimethoxy-7-hydroxyisoflavone	299.0919	299.0914	0.5	[M+H]+	298.0841	932.1259	870.8793	3.55	KJGPBYUQZLUKLL-UHFFFAOYSA-N	6.51	5.86	-0.65	0.043	Isoflavonoid, O-methylated
Biochanin A	285.0762	285.0758	0.4	[M+H]+	284.0685	881.004	911.9885	4.15	WUADCCWRTIWANL-UHFFFAOYSA-N	7.56	6.91	-0.65	0.154	Isoflavonoid, O-methylated

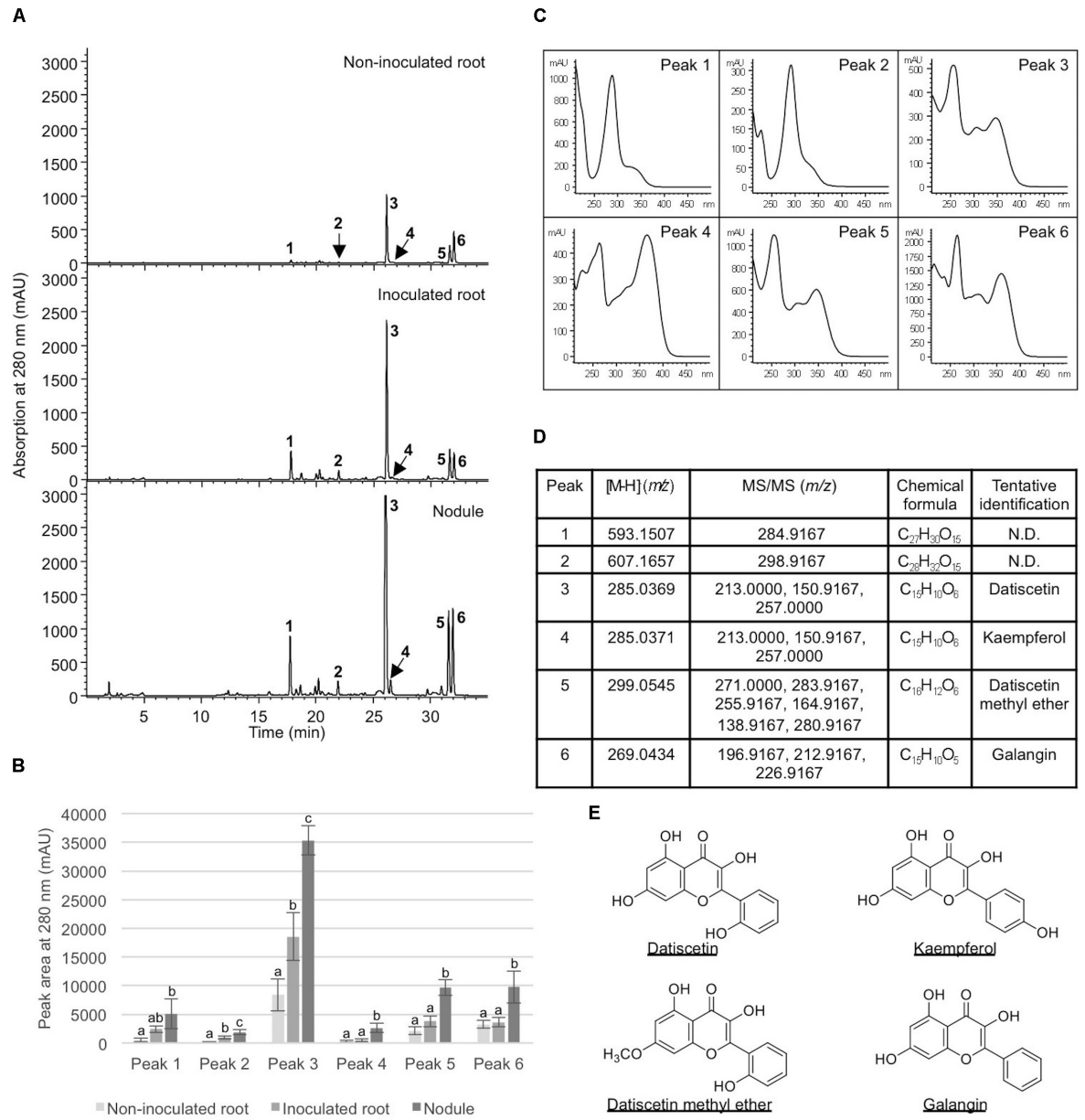
Scale	Log Root	Log Nodule	Log Fold Change
	10.00	10.00	2.00
	5.00	5.00	1.00
	0.00	0.00	0.00
			-1.00
			-2.00

FIGURE 4 | Flavonoids identified in *M. sativa* roots and nodules. All annotations conform to MSI level 2. For each flavonoid the abundance in both the roots and nodules is given as the log average peak height of four samples and presented as a heat-map. The names of molecules considered highly abundant are highlighted in blue. Significant changes in abundance between roots and nodules are highlighted in yellow.

Expression of Phenylpropanoid Pathway Genes in Transcriptomes of *D. glomerata* and *M. truncatula* Roots and Nodules

To understand the expression of the phenylpropanoid and flavonoid biosynthetic genes in roots and nodules, transcriptome data from roots and nodules of *D. glomerata* (Battenberg et al., 2018) and *M. truncatula* (Roux et al., 2014) were analyzed by differential expression between tissues and by comparing relative expression levels (as percentiles) of transcripts within each transcriptome. The reactions of the phenylpropanoid pathway, including the flavonoid branch, and the lignin-monolignol branch, are summarized in **Figure 1** and the relative expression of the genes comprising each branch are depicted in **Figure 6**. Early steps in the phenylpropanoid pathway preceding the flavonoid branch, e.g., phenylalanine ammonia-lyase (EC 4.3.1.24), and 4-coumarate-CoA ligase (4CL, EC 6.2.1.12) were highly expressed, above the 95th percentile in roots and nodules of both *D. glomerata* and *M. truncatula*, but were not significantly up-regulated or down-regulated in nodules of either host (**Figure 6**). The enzyme 4CL is responsible for the conversion of cinnamic acid to cinnamoyl-CoA and p-coumaric acid to p-coumaroyl-CoA, the precursor to flavonoid chalcones (**Figure 1**). In *D. glomerata* nodules, of the nine annotated 4CL transcripts, seven were not up-regulated relative to roots, while two (DgTrNR01535_a1_i1 and DgTrNR01535_a1_i2) were up-regulated over four-fold. In *M. truncatula* nodules, several annotated transcripts of this gene were also up-regulated between four- and six-fold, while one transcript was up-regulated greater than 300-fold (**Figure 6**).

Two of the four annotated transcripts of cinnamic acid 4-hydroxylase (C4H, EC 1.14.13.11), also known as *trans*-cinnamate 4-monooxygenase, were down-regulated in *D. glomerata* nodules almost 10-fold, while the other two were not statistically different (**Figure 6**) and none were expressed above the 87th percentile. C4H is the enzyme that catalyzes the last step in phenylpropanoid biosynthesis that precedes the separation of the naringenin flavonoid branch and the lignin-monolignol branch (**Figure 1**). In *M. truncatula*, by contrast, no C4H transcript was down-regulated in the nodule and one transcript (Medtr5g075450) was extremely highly expressed, above the 98th percentile of all genes in the transcriptome (**Figure 6**). The most highly expressed C4H transcripts expressed at significantly different levels in the



Peak	[M-H] (m/z)	MS/MS (m/z)	Chemical formula	Tentative identification
1	593.1507	284.9167	$C_{27}H_{30}O_{15}$	N.D.
2	607.1657	298.9167	$C_{28}H_{32}O_{15}$	N.D.
3	285.0369	213.0000, 150.9167, 257.0000	$C_{15}H_{10}O_6$	Datiscetin
4	285.0371	213.0000, 150.9167, 257.0000	$C_{15}H_{10}O_6$	Kaempferol
5	299.0545	271.0000, 283.9167, 255.9167, 164.9167, 138.9167, 280.9167	$C_{16}H_{12}O_6$	Datiscetin methyl ether
6	269.0434	196.9167, 212.9167, 226.9167	$C_{15}H_{10}O_5$	Galangin

FIGURE 5 | Analysis of metabolites in *Datisca glomerata* roots and nodules. **(A)** Representative HPLC elution profiles of phenolic metabolites extracted from non-inoculated and inoculated *D. glomerata* roots as well as nodules. Peaks (1–6) that show differential accumulation in roots and nodules are indicated. **(B)** Peak areas of differentially accumulated phenolic metabolites. Data shown are the average of four biological replicates ± standard deviation. Different letters indicate significant ($P < 0.05$) differences in metabolite levels among non-inoculated root, inoculated root, and nodule for each peak. **(C)** Absorption spectra of peaks 1–6. **(D)** MS and MS/MS analyses of peaks 1–6. **(E)** Chemical structures of tentatively identified phenolic metabolites in *D. glomerata* roots and nodules.

transcriptomes of *D. glomerata* and *M. truncatula* nodules ($p < 5 \times 10^{-4}$).

As shown in **Figure 4**, in the flavonoid branch, one transcript of chalcone isomerase (EC 5.5.1.6), which catalyzes the synthesis of flavonoids from chalcones, was highly up-regulated in *D. glomerata* nodules (approximately 50-fold) while the second was not significantly different. Of the 10 chalcone isomerase transcripts in *M. truncatula*, on the other hand, none were up-regulated in the nodule, and half of them were significantly down-regulated. Transcripts annotated as encoding flavanone 3-dioxygenase (EC number 1.14.13.21), the enzyme responsible for the synthesis of both eriodictyol from naringenin and dihydroquercetin from dihydrokaempferol (precursors to flavones and flavanols, respectively), were also among the most up-regulated flavonoid biosynthesis genes identified in the *D. glomerata* nodule (approximately 20- and 100-fold). By contrast, *M. truncatula* showed down-regulation of these genes in

the nodule or no significant difference between nodules and roots.

Four transcripts were annotated as flavonol synthase (flavonol biosynthesis, EC 1.14.11.23) in *D. glomerata*, one of which was significantly up-regulated in the nodule and was expressed in the 99th percentile in the nodule (**Figure 6**). In *M. truncatula* the corresponding transcripts were not significantly different between roots and nodules and were expressed in the 57th percentile at most, a marked contrast. Three flavonol-3-*O*-glucosyltransferases (EC 2.4.1.91) were annotated in *D. glomerata*; one of which showed high expression (above the 93rd percentile) with no significant change in expression between nodules and roots. In the transcriptome of *M. truncatula*, no transcripts were annotated as flavonol-3-*O*-glucosyltransferases.

Strikingly, no transcripts in the isoflavonoid pathway were annotated in the *D. glomerata* transcriptome, whereas, in the *M. truncatula* transcriptome, there were multiple transcripts annotated as encoding several enzymes in isoflavonoid biosynthesis (**Figure 6**). One of the eight transcripts of 2,7,4′-trihydroxyisoflavanone 4′-*O*-methyltransferase (EC 2.1.1.212) was highly expressed, above the 93rd percentile while four others were up-regulated over 10-fold in *M. truncatula* nodules. Enzymes that catalyze the synthesis of daidzein derivatives including formononetin and daidzein 7-*O*-glucoside were expressed but not significantly different between roots and nodules. Three transcripts of isoflavone 7-*O*-glucoside-6′′-*O*-malonyltransferase (EC 2.3.1.115), an enzyme known to synthesize malonated daidzein derivatives, were annotated in *M. truncatula*. One was down-regulated over 10-fold in the *M. truncatula* nodule relative to the roots, falling to around the 45th percentile in the nodule while the other two remained around the 72nd percentile.

The gene encoding the earliest enzyme in anthocyanin biosynthesis, anthocyanidin 3-*O*-glucosyltransferase (EC 2.4.1.115), was up-regulated more than 10-fold in *D. glomerata* nodules and two other transcripts were expressed above the 90th percentile (**Figure 6**). By contrast, no anthocyanidin 3-*O*-glucosyltransferase transcripts were annotated in the nodule or root transcriptomes of *M. truncatula*.

Genes encoding the major enzymes in the lignin-monolignol branch of the phenylpropanoid pathway showed generally low expression in the transcriptome of *D. glomerata* (**Figure 6**). The most highly expressed transcript of the first enzyme in this branch, shikimate *O*-hydroxycinnamoyltransferase (HCT, EC 2.3.1.133) was expressed in the 48th percentile. Coumaroyl-shikimate 3′-monooxygenase (EC 1.14.13.36), the next gene in the pathway, was expressed in the 61st percentile. Caffeoyl CoA-*O*-methyltransferase (EC 2.1.104), converting caffeoyl-CoA to feruloyl-CoA, was the exception. Two transcripts were expressed in the 90th percentile in *D. glomerata* nodules, although transcripts were expressed up to 10-fold higher in the roots. In *M. truncatula* nodules, the transcripts encoding HCT were expressed much more highly than HCT in *D. glomerata* ($p < 5 \times 10^{-5}$), around the 80th percentile. No transcripts of coumaroyl-shikimate 3′-monooxygenase were annotated in *M. truncatula*.

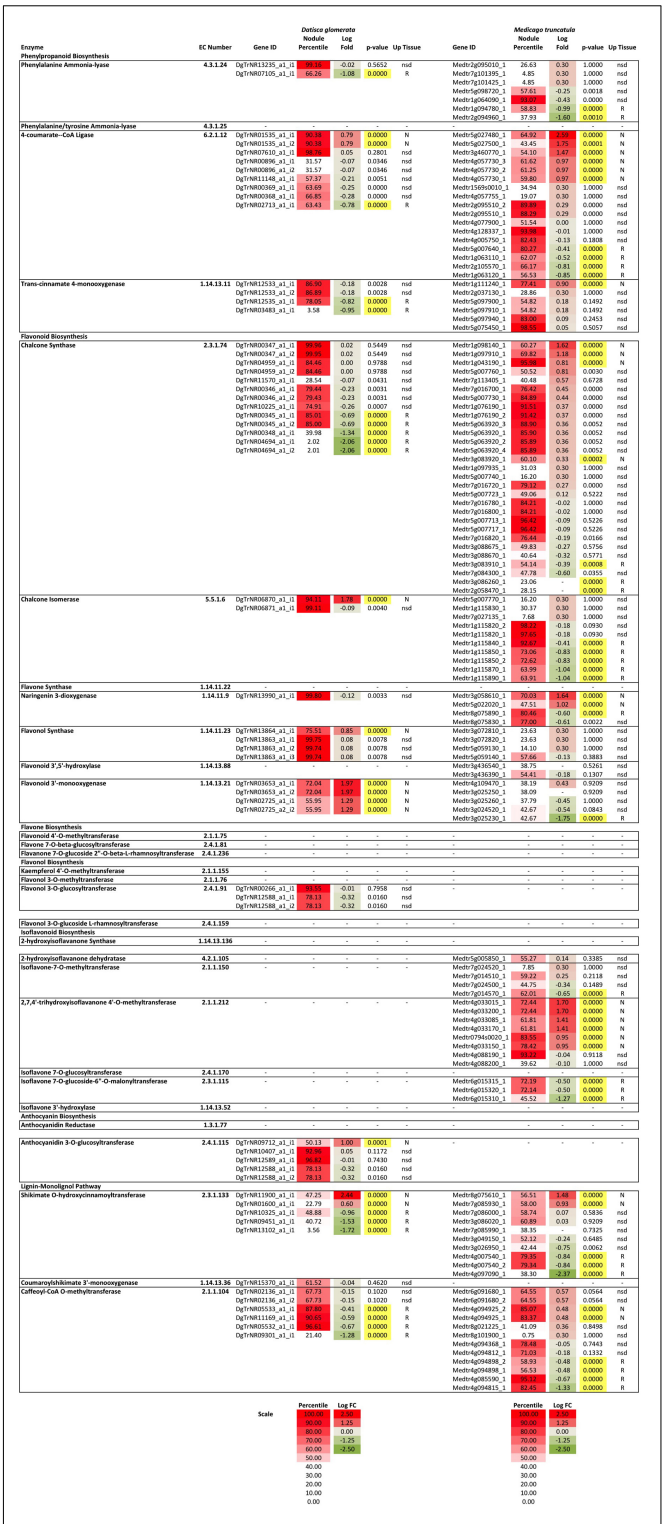


FIGURE 6 | Heat map of flavonoid biosynthesis gene expression in *D. glomerata* and *M. truncatula* including relative expression in the nodule transcriptome based on percentile of all genes in their respective transcriptome (Percentile) and fold change between nodules and roots (Log fold change). Significance level used for differences in expression between roots and nodules was $p < 0.001$. Up-regulated tissues: N, nodules; R, roots; and nsd, no significant difference.

DISCUSSION

Metabolite and Expression Analyses Provide New Insights Into Flavonoid Metabolism in *D. glomerata* Roots and Nodules

A range of flavonoids were synthesized in roots and nodules of *D. glomerata*, including flavones, flavonols, flavanones, anthocyanins, and isoflavonoids. A greater number of flavonoids were more abundant in the nodule relative to roots than vice versa (**Figure 3**), suggesting an overall increase in flavonoid biosynthesis in the nodule.

For several highly-abundant metabolites analyzed, there was a general trend of increasing concentration, when comparing uninoculated roots with either roots post-inoculation, or root nodules (**Figure 5**). This suggests that inoculation with *Frankia* induces some change in flavonoid metabolism in roots, either systemically in nodulated plants, or locally by association with *Frankia* in the rhizosphere. In a split-root experiment on *M. sativa,* initiation of nodulation by application of either the symbiont *Sinorhizobium meliloti* or its Nod factor to roots on one side of the plant led to increased amounts of daidzein on the uninoculated side (Catford et al., 2006), supporting an interpretation that flavonoid biosynthesis and distribution is under global regulation in RNS, similar to, or part of, autoregulation (Reid et al., 2011).

Derivatives of the pinocembrin-derived subclass of flavonoids, especially datiscetin and datiscin, represented some of the most abundant molecules in *D. glomerata* nodules (**Figure 3**). The biosynthesis of datiscetin has been proposed to proceed by a reaction similar to the synthesis of galangin from pinobanksin through the addition of a 2C–3C double bond to dihydrodatiscetin that is itself synthesized from pinobanksin (Grambow and Grisebach, 1971). Galangin biosynthesis utilizes flavonol synthase (EC 1.14.11.23), which was found to be up-regulated four-fold in *D. glomerata* nodules (**Figure 6**). Because flavonol synthase has been shown to catalyze multiple reactions including both kaempferol and galangin biosynthesis (Miyahisa et al., 2006), it seems likely that datiscetin biosynthesis could be performed by this enzyme as well. Dihydrodatiscetin itself, however, was not identified in roots or nodules of *D. glomerata* in this study (**Figure 3**).

The prevalence of pinocembrin and its derivatives in *D. glomerata* is likely directly related to the relatively low expression of C4H (EC 1.14.13.11) (**Figure 5**), which appears to be a pivotal enzyme in the phenylpropanoid pathway in *D. glomerata* whose expression impacts two separate branch points (**Figure 1**): first, the enzyme catalyzes the conversion of the pinocembrin-precursor cinnamic acid to naringenin-precursor p-coumaric acid, and thus controls the balance between the pinocembrin and naringenin flavonoid branches. Because the early enzymes in the flavonoid branch, including chalcone synthase (EC 2.3.1.74) and naringenin 3-diooxygenase (EC 1.14.11.9) are multi-functional, catalyzing reactions with multiple substrates (Martens et al., 2010), the altered flux favoring cinnamoyl-CoA in *D. glomerata* likely directs the flow of flavonoid biosynthesis more toward pinocembrin and ultimately datiscetin and datiscin. Second, because expression of C4H is diminished relative to *M. truncatula,* the synthesis of naringenin-based flavonoids is likely aided in *D. glomerata* by lower expression of HCT (EC 2.3.1.133), which decreases metabolic flux to lignins in favor of flavonoid biosynthesis, a pattern similar to what was shown to occur when HCT was down-regulated in *M. sativa* by Gallego-Giraldo et al. (2014).

C4H has been shown to be the major rate-limiting step in lignin biosynthesis (Anterola and Lewis, 2002) suggesting that it also functions to control the relative flux of phenylpropanoids between flavonoid and lignin biosynthesis. Control of flux between flavonoid and lignin branches of the phenylpropanoid pathway in plants could be useful for crop improvement to enhance flavonoid content, particularly the pinocembrin pathway. Flavonoids in general are nutritional antioxidants, and pinocembrin specifically has shown both antitumor and neuroprotective capabilities (Rasul et al., 2013).

In addition to datiscetin, a flavonol synthesized from pinocembrin as discussed above, other flavonols were abundant in *D. glomerata* roots and nodules, particularly quercetin and its glycosides (**Figure 3**). Flavonol synthase (EC 1.14.11.23) was very highly expressed in *D. glomerata* roots and nodules, with one transcript expressed above the 99th percentile, compared to 57th percentile at the highest in *M. truncatula* nodules; and the other transcript was significantly up-regulated in the *D. glomerata* nodule seven-fold (**Figure 6**), reflecting the great abundance of flavonols, which may perform a variety of roles in nodules. Quercetin has been shown to regulate auxin gradients in roots during nodule formation as measured by a *gusA* gene-auxin response promoter construct (Mathesius et al., 1998). Quercetin showed a higher level of auxin transport inhibiting activity than kaempferol, apigenin, naringenin, or genistein; glycosylation decreased the auxin-inhibitory effect (Mathesius et al., 1998, 2015). Interestingly, in the laser-capture microdissection study of developmental gene expression in *M. truncatula* performed by Roux et al. (2014), 100% of flavonol synthase transcripts (Medtr5g059140) in nodules were found in zone FIID, the second distal fraction of the nodule. Cells in this zone are undergoing expansion and rhizobial infection (Roux et al., 2014), making it the zone where auxin gradient inhibition is likely required during the nodulation process (Mathesius et al., 2015). Additionally, quercetin has been shown to arrest cell division and cause DNA breaks leading to endoreduplication in eukaryotic cells *in vitro* (Cantero et al., 2006). This may suggest another potential role for flavonols in the formation of symbiotic cells in nodules (Vinardell et al., 2003), in addition to CCS52A-mediated endoreduplication, as described by Adachi et al. (2011). Finally, flavonols have been shown to protect enzymes from deactivation by nitric oxide and peroxynitrite (Heijnen et al., 2001). Nitric oxide is a signaling molecule required for nodule formation, however, it also deactivates enzymes important for symbiosis, including nitrogenase and glutamine synthetase, by nitration of tyrosine residues (Melo et al., 2011). Flavonoids in general have been found to protect enzymes by scavenging nitric oxide and peroxynitrite. Heijnen et al. (2001), found that galangin had strong scavenging activity and correlated this activity with the

hydroxyl group on the third carbon; this would suggest that datiscetin is a strong peroxynitrite scavenger as well.

A feature of the flavonoid glycosides in *D. glomerata* roots and nodules was the frequent occurrence of rutinose glycosides. Six flavonoids were identified with rutinose glycosylations. Rutinose is a common glycosylation of flavonoids in many plants (Cuyckens et al., 2001), and was previously shown to occur in leaves of *Datisca cannabina* (Bohm, 1988). Rutinose was reported to occur as an exceptionally abundant free sugar in roots, nodules, and leaves of *D. glomerata* and *D. cannabina* (Schubert et al., 2010) that could not be hydrolyzed in cell extracts. Thus it is hypothesized that rutinose is synthesized as a flavonoid glycosylation and then released as a free sugar (Schubert et al., 2010).

Potential Nod Gene Inducing Flavonoids in *Datisca glomerata*

In legume symbioses a range of flavonoids, predominantly aglycones, have been shown to induce the expression of *nod* genes in rhizobia, including flavones (luteolin), flavanones (eriodictyol and naringenin), and chalcones in *Medicago* and isoflavonoids (daidzein and genistein) in *Glycine* (Phillips and Tsai, 1992). In this study all of these metabolites were identified in roots and nodules of both *D. glomerata* and *M. sativa*, with the exception of chalcones, which were not detected in either host (**Figures 3**, **4**). The identified molecules could play potentially similar roles in the *D. glomerata* symbiosis as in *Medicago*. *D. glomerata* is nodulated by Cluster 2 *Frankia* strains whose genomes have been shown to contain *nodABC* gene homologs that are expressed in symbiosis (Persson et al., 2015; Nguyen et al., 2016); however, it is unknown whether flavonoids are involved in the regulation of these genes since no clear *nodD* homolog has been identified (Persson et al., 2015). Our results indicate that molecules in the pinocembrin pathway, including galangin and datiscetin, are possible candidates for a similar role in *D. glomerata*, because they are synthesized in great abundance (**Figure 3**) and are synthesized by uninoculated roots as well as inoculated roots and nodules (**Figure 5**) suggesting they are present in roots before the roots are infected.

Pinocembrin, the precursor to galangin and datiscetin, has been reported in several actinorhizal hosts. In the Fagales, it was found in the leaves and flowers of members of the genus *Alnus* (Betulaceae; Ren et al., 2017), as well as in leaves of *Myrica* and *Comptonia* (Myricaceae), all hosts nodulated by *Frankia* belonging to Cluster 1 (Wollenweber et al., 1985; Normand et al., 1996). However, flavonoids from leaf exudates of eight species of *Ceanothus* (Rhamnaceae, in Rosales), another genus that, like *D. glomerata,* is nodulated by Cluster 2 *Frankia* (Normand et al., 1996), were not reported to include pinocembrin or its derivatives (Wollenweber et al., 2004).

D. glomerata and *M. sativa* Share Similar Classes of Flavonoids but Differ in Abundance

Both *D. glomerata* and *M. sativa* metabolomes contained flavonoids in a range of classes. Both hosts contained similar classes of flavonoids, including flavanones, flavonols, flavonoid glycosides, isoflavonoids, and isoflavonoid glycosides, but within each class they varied significantly in diversity (**Figure 2**). In both hosts the abundance of the majority of flavonoids was not statistically different between their roots and nodules.

The largest differences between the two plants were in the amount within particular flavonoid classes produced, most notably, in the pinocembrin pathway as discussed above, and in the isoflavonoids. All four of the highly abundant flavonoids in *M. sativa* nodules were isoflavonoids (**Figure 4**). This conforms to earlier reports highlighting isoflavonoids as most abundant in nodules of *M. sativa* (Tiller et al., 1994) and *M. truncatula* (Modolo et al., 2007; Staszkow et al., 2011). In *D. glomerata*, however, only one of the 10 highly abundant flavonoids (daidzein) was an isoflavonoid (**Figure 3**). Isoflavonoids identified in *D. glomerata* were found early in the biosynthetic pathway (**Figure 3**) whereas *M. sativa* included highly abundant derived isoflavonoids, including the pterocarpin medicarpin and its precursor formononetin (**Figure 4**). Pterocarpins are primarily found in legumes (Dewick, 1982) and have been shown to function as antimicrobials that show much greater inhibition of gram-positive bacteria than gram-negatives (Gnanamanickam and Smith, 1980). However, Auguy et al. (2011) reported expression of isoflavone reductase in the actinorhizal plant *C. glauca*, suggesting the synthesis of the more derived isoflavonoids similar to the *Medicago*. This leaves the distribution and role of isoflavonoids in actinorhizal plants unresolved.

CONCLUSION

We present the first comparison of metabolic profiles of flavonoids from both roots and nodules of two host plants within the NFC, *D. glomerata* and *M. sativa,* with transcriptomes obtained from roots and nodules, in the context of phenylpropanoid biosynthetic pathways. The most abundant flavonoids in *D. glomerata* were derivatives of pinocembrin as well as naringenin whereas flavonoids from *M. sativa* were isoflavonoids and derivatives of naringenin. These findings correlate with the pattern of expression of cinnamic acid 4-hydroxylase (C4H), in the transcriptomes of the two hosts. *D. glomerata* showed relatively low expression of C4H in nodules compared to *M. truncatula*, suggesting a role for this enzyme in directing the flow of the phenylpropanoid pathway between the pinocembrin branch and the naringenin branch. Similarly, shikimate *O*-hydroxycinnamoyltrasferase (HCT), the link between the flavonoid and monolignol branches of the phenylpropanoid pathway, also showed lower expression in *D. glomerata*, supporting a difference in metabolic flux between the two hosts that favors flavonoids over monolignol/lignin production in *D. glomerata*.

Flavonoids of the same classes were present in roots and nodules of both *D. glomerata* and *M. sativa*, including flavanones, flavonols, and isoflavonoids, suggesting similar roles for flavonoids during nodule development and symbiosis across lineages in the NFC. Common roles may include

symbiotic signaling, protection of enzymes from nitration, nodule organogenesis including phytohormone regulation, and cell-cycle modification. To identify symbiotically important flavonoids, further higher resolution transcriptome studies including spatio-temporal sampling (as in Roux et al., 2014, or Larrainzar et al., 2015) in combination with metabolomics profiling are needed. Secondly, responses of *Frankia* in culture to purified flavonoids identified as unique or amplified in their respective hosts should be measured at the transcriptomic level, and in terms of nodulation patterns, to evaluate a broader role for flavonoids as signaling molecules in the actinorhizal symbioses.

DATA AVAILABILITY STATEMENTS

Metabolome data from *D. glomerata* and *M. sativa* generated in this study are presented in **Supplementary Table S2**. Transcriptome data for *D. glomerata* and *M. truncatula* used were obtained from Roux et al. (2014) and Battenberg et al. (2018), respectively.

AUTHOR CONTRIBUTIONS

IG developed metabolomic and transcriptomic profiles, determined patterns of flavonoid metabolism, and substantially wrote the manuscript; each author contributed to the writing of the manuscript for their, respective, section. In addition, KB provided annotated transcriptomes and was responsible for the plant material. AV developed analytical methods for and performed metabolomic analyses. AW performed LC-MS (MS/MS) analyses on peaks identified by HPLC and analyzed glycosyltransferase transcriptome data. LT carried out HPLC analyses and provided data interpretation. OF contributed to metabolomics project design and manuscript editing. AB provided project oversight and contributed to manuscript construction.

FUNDING

IG was supported in part by a UC Davis Department of Plant Sciences graduate student research assistantship.

ACKNOWLEDGMENTS

We appreciate technical assistance of undergraduate intern Jannah A. Wren. Metabolomic processing was performed in the West Coast Metabolomics Center, University of California, Davis, U24 DK097154 and U2C ES030158. The project was part of USDA NIFA CA-D-PLS-2173-H (AB).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2018.01463/full#supplementary-material

FIGURE S1 | Head-to-tail comparisons of MS/MS spectra of the most abundant flavonoids identified in *D. glomerata*. Reference library spectra are shown in red, experimental spectra are given in blue. Metadata include the similarity dot product and reverse dot product scores, InChIKey, and the structure of annotated compounds.

FIGURE S2 | Head-to-tail comparisons of MS/MS spectra of the most abundant flavonoids identified in *M. sativa*. Reference library spectra are shown in red, experimental spectra are given in blue. Metadata include the similarity dot product and reverse dot product scores, InChIKey, and the structure of annotated compounds.

TABLE S1 | Collection data for root and nodule samples of *D. glomerata* and *M. sativa*.

TABLE S2 | Flavonoid metabolome data obtained from LCMS analyses of *D. glomerata* and *M. sativa* nodules and roots.

TABLE S3 | Number of flavonoids annotated in each class in *D. glomerata* and *M. sativa* roots and nodules and chi-square test comparing the distribution of classes by proportion of total flavonoids in each host.

REFERENCES

Adachi, S., Minamisawa, K., Okushima, Y., Inagaki, S., Yoshiyama, K., Kondou, Y., et al. (2011). Programmed induction of endoreduplication by DNA double-strand breaks in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10004–10009. doi: 10.1073/pnas.1103584108

Anterola, A. M., and Lewis, N. G. (2002). Trends in lignin modification: a comprehensive analysis of the effects of genetic manipulations/mutations on lignification and vascular integrity. *Phytochemistry* 61, 221–294. doi: 10.1016/S0031-9422(02)00211-X

Auguy, F., Abdel-Lateif, K., Doumas, P., Badin, P., Guerin, B., Bogusz, D., et al. (2011). Activation of the isoflavonoid pathway in actinorhizal symbioses. *Funct. Plant Biol.* 38, 690–696. doi: 10.1071/FP11014

Battenberg, K., Potter, D., Tabuloc, C., Chiu, J. C., and Berry, A. M. (2018). Comparative transcriptomics of two actinorhizal plants and the legume Medicago truncatula support the homology of root nodule symbioses and is congruent with a two-step process of evolution in the nitrogen-fixing clade of angiosperms. *Front. Plant Sci.* doi: 10.3389/fpls.2018.01256

Bohm, B. A. (1988). Flavonoid systematics of the datiscaceae. *Biochem. Syst. Ecol.* 16, 151–155. doi: 10.1016/0305-1978(88)90088-9

Buer, C. S., Imin, N., and Djordjevic, M. A. (2010). Flavonoids: new roles for old molecules. *J. Integr. Plant Biol.* 52, 98–111. doi: 10.1111/j.1744-7909.2010.00905.x

Cajka, T., and Fiehn, O. (2015). Toward merging untargeted and targeted methods in mass spectrometry-based metabolomics and lipidomics. *Anal. Chem.* 88, 524–545. doi: 10.1021/acs.analchem.5b04491

Cantero, G., Campanella, C., Mateos, S., and Cortés, F. (2006). Topoisomerase II inhibition and high yield of endoreduplication induced by the flavonoids luteolin and quercetin. *Mutagenesis* 21, 321–325. doi: 10.1093/mutage/gel033

Catford, J. G., Staehelin, C., Larose, G., Piche, Y., and Vierheilig, H. (2006). Systemically suppressed isoflavonoids and their stimulating effects on nodulation and mycorrhization in alfalfa split-root systems. *Plant Soil* 285, 257–266. doi: 10.1007/s11104-006-9012-8

Champion, A., Lucas, M., Tromas, A., Vaissayre, V., Crabos, A., Diédhiou, I., et al. (2015). Inhibition of auxin signaling in *Frankia* species-infected cells in *Casuarina glauca* nodules leads to increased nodulation. *Plant Physiol.* 167, 1149–1157. doi: 10.1104/pp.114.255307

Cuyckens, F., Rozenberg, R., de Hoffman, E., and Claeys, M. (2001). Structure characterization of flavonoid O-diglycosides by positive and negative nano-electrospray ionization ion trap mass spectrometry. *J. Mass Spectrom.* 36, 1203–1210. doi: 10.1002/jms.224

DeFelice, B. C., Mehta, S. S., Samra, S., Čajka, T., Wancewicz, B., Fahrmann, J. F., et al. (2017). mass spectral feature list optimizer (MS-FLO): a tool to minimize false positive peak reports in untargeted liquid chromatography-mass spectroscopy (LC-MS) data processing. *Anal. Chem.* 89, 3250–3255. doi: 10.1021/acs.analchem.6b04372

Dewick, P. M. (1982). "Isoflavonoids," in *The Flavonoids: Advances in Research*, eds J. B. Harborne and T. J. Mabry (Cambridge: Cambridge University Press), 535–640. doi: 10.1007/978-1-4899-2915-0_10

Dunn, W. B., Erban, A., Weber, R. J. M., Creek, D. J., Brown, M., Breitling, R., et al. (2012). Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics* 9, 44–66. doi: 10.1007/s11306-012-0434-4

Feunang, Y. D., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G., et al. (2016). ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J. Cheminform.* 8, 1–20.

Gage, D. J. (2004). Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiol. Mol. Biol. Rev.* 68, 280–300. doi: 10.1128/MMBR.68.2.280-300.2004

Gallego-Giraldo, L., Bhattarai, K., Pislariu, C. I., Nakashima, J., Jikumaru, Y., Kamiya, Y., et al. (2014). Lignin modification leads to increased nodule numbers in alfalfa. *Plant Physiol.* 164, 1139–1150. doi: 10.1104/pp.113.232421

Gnanamanickam, S. S., and Smith, D. A. (1980). Selective toxicity of isoflavonoid phytoalexins to gram-positive bacteria. *Phytopathology* 70, 894–896. doi: 10.1094/Phyto-70-894

Grambow, H. K., and Grisebach, H. (1971). Further studies on the biosynthesis of flavonoids in *Datisca cannabina*. *Phytochemistry* 10, 789–796. doi: 10.1016/S0031-9422(00)97148-6

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084

Heijnen, C. G., Haene, G. R., van Acker, F. A., van der Vijgh, W. J., and Bast, A. (2001). Flavonoids as peroxynitrite scavengers: the role of the hydroxyl groups. *Toxicol. Vitro* 15, 3–6. doi: 10.1016/S0887-2333(00)00053-9

Hoagland, D. R., and Arnon, D. I. (1950). The water-culture method for growing plants without soil. *Calif. Agric. Exp. Stn. Circ.* 347.

Jones, P., Binns, D., Chang, H., Fraser, M., Li, W., AcAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070

Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., et al. (2016). PubChem substance and compound databases. *Nucleic Acids Res.* 44, D1202–D1213. doi: 10.1093/nar/gkv951

Knollenberg, B. J., Liu, J., Yu, S., Lin, H., and Tian, L. (2018). Cloning and functional characterization of a p-coumaroyl quinate/shikimate 3'-hydroxylase from potato (Solanum tuberosum). *Biochem. Biophys. Res. Commun.* 496, 462–467. doi: 10.1016/j.bbrc.2018.01.075

Larrainzar, E., Riely, B., Kim, S. C., Carrasquilla-Garcia, N., Yu, H., Hwang, H., et al. (2015). Deep sequencing of the *Medicago truncatula* root transcriptome reveals a massive and early interaction between Nod factor and ethylene signals. *Plant Physiol.* 169, 233–265. doi: 10.1104/pp.15.00350

Long, S. R. (1996). Rhizobium symbiosis: nod factors in perspective. *Plant Cell* 8, 1885–1898. doi: 10.1105/tpc.8.10.1885

Ma, Y., Tanaka, N., Vaniya, A., Kind, T., and Fiehn, O. (2016). Ultrafast polyphenol metabolomics of red wines using MicroLC-MS/MS. *J. Agric. Food Chem.* 64, 505–512. doi: 10.1021/acs.jafc.5b04890

Markmann, K., and Parniske, M. (2009). Evolution of root endosymbiosis with bacteria: how novel are nodules? *Trends Plant Sci.* 14, 77–86. doi: 10.1016/j.tplants.2008.11.009

Martens, S., Preub, A., and Matern, U. (2010). Multifunctional flavonoid dioxygenases: flavonol and anthocyanin biosynthesis in *Arabidopsis thaliana* L. *Phytochemistry* 71, 1040–1049. doi: 10.1016/j.phytochem.2010.04.016

Mathesius, U., Jin, J., van Noorden, G. E., Ng, P. J., and Wasson, A. P. (2015). "Regulation of nodule development by short- and long-distance auxin transport control," in *Biological Nitrogen Fixation*, Vol. 1, ed. F. J. de Bruijn (Hoboken, NJ: John Wiley & Sons), 465–473.

Mathesius, U., Schlaman, H. R. M., Spaink, H. P., Sautter, C., Rolfe, B. G., and Djordjevic, M. A. (1998). Auxin transport inhibition precedes root nodule formation in white clover roots and is regulated by flavonoids and derivatives of chitin oligosaccharides. *Plant J.* 14, 23–34. doi: 10.1046/j.1365-313X.1998.00090.x

Melo, P. M., Silva, L. S., Ribeiro, I., Seabra, A. R., and Carvalho, H. G. (2011). Glutamine synthetase is a molecular target of nitric oxide in root nodules of *Medicago truncatula* and is regulated by tyrosine nitration. *Plant Physiol.* 157, 1505–1517. doi: 10.1104/pp.111.186056

Miyahisa, I., Funa, N., Ohnishi, Y., Martens, S., Moriguchi, T., and Horinouchi, S. (2006). Combinatorial biosynthesis of flavones and flavonols in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 71, 53–58. doi: 10.1007/s00253-005-0116-5

Modolo, L. V., Blount, J. W., Achnine, L., Naoumkina, M. A., Wang, X., and Dixon, R. A. (2007). A functional genomics approach to (iso)flavonoid glycosylation in the model legume *Medicago truncatula*. *Plant Mol. Biol.* 64, 499–518. doi: 10.1007/s11103-007-9167-6

Neumann, S., and Bocker, S. (2010). Computational mass spectrometry for metabolomics: identification of metabolites and small molecules. *Anal. Bioanal. Chem.* 398, 2779–2788. doi: 10.1007/s00216-010-4142-5

Nguyen, T. V., Wibberg, D., Battenberg, K., Blom, J., Vanden Heuvel, B., Berry, A. M., et al. (2016). An assemblage of *Frankia* cluster II strains from California contains the canonical nod genes and also the sulfotransferase gene nodH. *BMC Genomics* 17:796. doi: 10.1186/s12864-016-3140-1

Normand, P., Orso, S., Cournoyer, B., Jeannin, P., Chapelon, C., Dawson, J., et al. (1996). Molecular phylogeny of the genus Frankia and related genera and emendation of the family *Frankiaceae*. *Int. J. Syst. Evol. Microbiol.* 46, 1–9.

Oldroyd, G. E. D. (2013). Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat. Rev. Microbiol.* 11, 252–263. doi: 10.1038/nrmicro2990

Ono, N., Qin, X., Wilson, A., Li, G., and Tian, L. (2016). Two UGT84 family glycosyltransferases catalyze a critical reaction of hydrolyzable tannin biosynthesis in pomegranate (*Punica granatum*). *PLoS One* 11:e0156319. doi: 10.1371/journal.pone.0156319

Peck, M. C., Fisher, R. F., and Long, S. R. (2006). Diverse flavonoids stimulate NodD1 binding to nod gene promoters in *Sinorhizobium meliloti*. *J. Bacteriol.* 188, 5417–5427. doi: 10.1128/JB.00376-06

Péret, B., Swarup, R., Jansen, L., Devos, G., Auguy, F., Collin, M., et al. (2007). Auxin influx activity is associated with *Frankia* infection during actinorhizal nodule formation in *Casuarina glauca*. *Plant Physiol.* 144, 1852–1862. doi: 10.1104/pp.107.101337

Persson, T., Battenberg, K., Demina, I. V., Vigil-Stenman, T., Vanden Heuvel, B., Pujic, P., et al. (2015). Candidatus Frankia datiscae Dg1, the actinobacterial microsymbiont of *Datisca glomerata*, expresses the canonical nod genes nodABC in symbiosis with its host plant. *PLoS One* 10:e0127630. doi: 10.1371/journal.pone.0127630

Phillips, D. A., and Tsai, S. M. (1992). Flavonoids as plant signals to rhizosphere microbes. *Mycorrhiza* 1, 55–58. doi: 10.1007/BF00206136

Rasul, A., Millimouno, F. M., Wltayb, W. A., Ali, M., Li, J., and Li, X. (2013). Pinocembrin: a novel natural compound with versatile pharmacological and biological activities. *Biomed. Res. Int.* 2013:379850. doi: 10.1155/2013/379850

Reid, D. E., Ferguson, B. J., Hayashi, S., Lin, Y., and Gresshoff, P. M. (2011). Molecular mechanisms controlling legume autoregulation of nodulation. *Ann. Bot.* 108, 789–795. doi: 10.1093/aob/mcr205

Ren, X., He, T., Chang, Y., Zhao, Y., Chen, X., Bai, S., et al. (2017). The genus Alnus, a comprehensive outline of its chemical constituents and biological activities. *Molecules* 22:e1383. doi: 10.3390/molecules22081383

Roux, B., Rodde, N., Jardinaud, M., Timmers, T., Sauviac, L., Cottret, L., et al. (2014). An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant J.* 77, 817–837. doi: 10.1111/tpj.12442

Schubert, M., Melnikova, A. N., Mesecke, N., Zubkova, E. K., Fortte, R., Batashev, D. R., et al. (2010). Two novel disaccharides, rutinose and methylrutinose, are involved in carbon metabolism in *Datisca glomerata*. *Planta* 231, 507–521. doi: 10.1007/s00425-009-1049-5

Schymanski, E. L., Jeon, J., Guide, R., Fenner, K., Ruff, M., Singer, H. P., et al. (2014). Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ. Sci. Technol.* 48, 2097–2098. doi: 10.1021/es5002105

Shirley, B. W. (1996). Flavonoid biosynthesis: 'new' functions for an 'old' pathway. *Trends Plant Sci.* 1, 377–382.

Soltis, D. E., Soltis, P. S., Morgan, D. R., Swensen, S. M., Mullin, B. C., Dowd, J. M., et al. (1995). Chloroplast gene sequence data suggest a single orgin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proc. Natl. Acad. Sci. U.S.A.* 92, 2647–2651. doi: 10.1073/pnas.92.7.2647

Staszkow, A., Swarcewicz, B., Banasiak, J., Muth, D., Jasinski, M., and Stobiecki, M. (2011). LC/MS profiling of flavonoid glycoconjugates isolated from hairy roots, suspension root cell cultures and seedling roots of Medicago truncatula. *Metabolomics* 7, 604–613. doi: 10.1007/s11306-011-0287-2

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., et al. (2007). Proposed minimum reporting standards for chemical analysis. *Metabolomics* 3, 211–221. doi: 10.1007/s11306-007-0082-2

Tiller, S. A., Parry, A. D., and Edwards, R. (1994). Changes in the accumulation of flavonoid and isoflavonoid conjugates associated with plant age and nodulation in alfalfa (*Medicago sativa*). *Physiol. Plant.* 91, 27–36. doi: 10.1111/j.1399-3054.1994.tb00655.x

Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., et al. (2015). MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* 12, 523–526. doi: 10.1038/nmeth.3393

Vinardell, J. M., Fedorova, E., Cebolla, A., Kevei, Z., Horvath, G., Kelemen, Z., et al. (2003). Endoreduplication mediated by the anaphase-promoting complex activator CCS523A is required for symbiotic cell differentiation in *Medicago truncatula* nodules. *Plant Cell* 15, 2093–2105. doi: 10.1105/tpc.014373

Wasson, A. P., Pellerone, F. I., and Mathesius, U. (2006). Silencing the flavonoid pathway in *Medicago truncatula* inhibits root nodule formation and prevents auxin transport regulation by rhizobia. *Plant Cell* 18, 1617–1629. doi: 10.1105/tpc.105.038232

Wollenweber, E., Dorr, M., Bohm, B. A., and Roitman, J. N. (2004). Exudate flavonoids of eight species of *Ceanothus* (Rhamnaceae). *Z. Naturforsch.* 59, 459–462. doi: 10.1515/znc-2004-7-801

Wollenweber, E., Kohorst, G., Mann, K., and Bell, J. M. (1985). Leaf gland flavonoids in *Comptonia peregrina* and *Myrica pensylvanic* (Myricaceae). *J. Plant Physiol.* 117, 423. doi: 10.1016/S0176-1617(85)80049-3

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Gifford, Battenberg, Vaniya, Wilson, Tian, Fiehn and Berry. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.